IRE Transactions

on INFORMATION THEORY

Volume IT-3 MARCH, 1957 Number 1

In This Issue

For complete Table of Contents, see page 1.

PUBLISHED BY THE
Professional Group on Information Theory

IRE PROFESSIONAL GROUP ON INFORMATION THEORY

The Professional Group on Information Theory is an organization, within the framework of the IRE, of members with principal professional interest in Information Theory. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of the prescribed annual assessment of $2.00.

IRE TRANSACTIONS®
on Information Theory

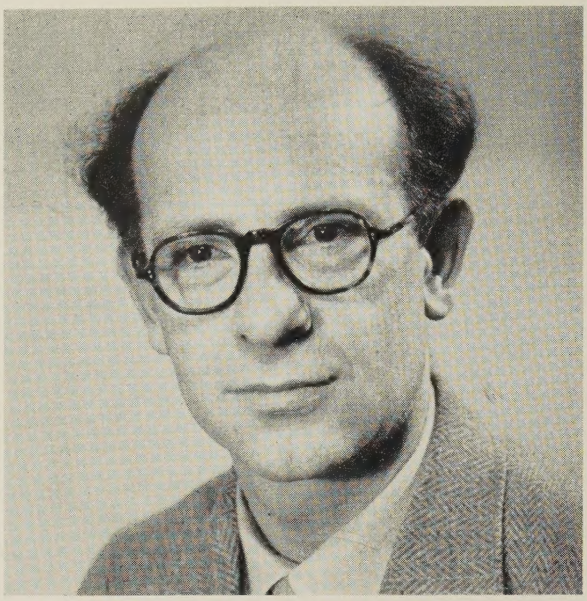
Published by the Institute of Radio Engineers, Inc., for the Professional Group on Information Theory at 1 East 79th Street, New York 21, N. Y. Responsibility for the contents rests upon the authors, and not upon the Institute, the Group or its members. Single copy prices: IRE-PGIT members, $2.20; IRE members, $3.30; nonmembers, $6.60.

IRE Transactions on Information Theory

Published Quarterly by the Professional Group on Information Theory

Volume IT-3 March, 1957 Number 1

TABLE OF CONTENTS



Philip M. Woodward

Philip M. Woodward is a mathematician at the Radar Research Establishment, England. Born in Staffordshire, he was educated at Blundell's School and Wadham College, Oxford, where he was a Methuen Scholar in mathematics. After graduation in 1941, he joined "T.R.E." (now R.R.E.) and during the war years worked on radio propagation under Dr. Henry G. Booker.

He has written various papers on antenna theory, noise theory, and computing. In the field of information theory (broadly defined), his work with I. L. Davies was an attempt to express radar problems in an absolute form, and this led to the publication in 1953 of a book entitled "Probability and Information Theory with Applications to Radar." In 1956, he visited the United States as a Gordon Mackay Visiting Lecturer at Harvard University. His present interest lies in the development of automatic programming for electronic computers.

Mr. Woodward is a Chartered Electrical Engineer, a member of the Ratio Club, and for recreation an amateur clockmaker and harpsichordist.

5856-129

Entropy and Negentropy

PHILIP M. WOODWARD

Many of us are not physicists. We are radio engineers and applied mathematicians, but our work borders so closely on physics that we take a particular interest in the reaction of the physicists to our special topic. It has been a rather unfortunate one. Certainly there is a relationship between the mathematical theory of communication and statistical thermodynamics, if only because both employ a familiar formula.

$$X = -\sum p_i \log p_i$$

to represent a certain fundamental quantity. This quantity springs naturally from what we might call "average logarithmic counting"—of different possible messages in communication theory and quantum states in physics. The coincidence, if such it can be called, was noted from the start both by Wiener and by Shannon, and it gave rise to a widespread feeling that the two subjects ought in some way to coalesce. But whereas X appears in both in chapter one, a purposeful divergence takes place in chapter 2 because the aims are quite different. Interaction between information theory and statistical physics has thus been limited to little more than an interchange of terms. Yet even in this there have been some widespread misconceptions which we ought to try finally to clear up.

First, what are the facts? Communication theorists have borrowed the term entropy from the physicists to describe the quantity X, and following Shannon they use the letter H for it. Physicists use the letter S, and their units are different, so we have

Shannon's $H = X$ (dimensionless entropy)

Physicists' $S = kX$ (physical entropy).

But the letter H also figures in Boltzmann's H-Theorem where, in spite of a remark to the contrary in Shannon's original paper, we find that

Boltzmann's $H = -X$ (negative dimensionless entropy).

Although physicists have often felt some uneasiness about the sign of H in information theory, this particular clash of notation does not appear to have been the cause of it. Rather it was the use of a single quanity to measure simultaneously uncertainty and information which originally gave rise to difficulty. An explanation was given in 1952 by D. K. C. MacDonald writing in the *Journal of Applied Physics*, and the last word—so we thought—had been said. But recently an eminent physicist, Leon Brillouin, has remarked that Shannon's H is *negative* entropy because it is reduced and not increased by the action of an irreversible filter. It seems to have been a hasty conclusion, and the probable explanation of the puzzle is this. In the communication problem, we are interested only in the *electrical* form of the signals and not in the heat dissipated by them when they flow through resistors in filters. The physicist must include this disorganized radiation, and his total entropy will then be found to increase without his having to change any definitions.

Whatever the true explanation may be, it should be made clear without any doubt or hesitation that the term entropy is used in information theory by mathematical analogy. The expression X, which we call the entropy of a set of probabilities, is identical in form *and in sign* with the expression for entropy in physics. The units may differ, and the probabilities stand for different things, but neither of these can change the sign of X. The pure information theorist, of course, is left unmoved by arguments about mere names. In a mathematical theory, definitions are judged solely in the light of the theorems to which they lead, and word-names are arbitrary, even irrelevant. But when we apply the theory, or compare two different theories, it is surely important to avoid using one name for two different things or two opposite names for one and and the same thing.

The IRE "Affiliate" Plan—A New Venture in Engineering Society Structure and Service

W. R. G. BAKER, *Chairman*
IRE Professional Groups Committee

On January 4, 1957, the IRE Board of Directors arrived at a decision which may in time prove to be one of the most far-reaching in its 45-year history. On that date the Board adopted a plan which will enable non-IRE members whose main professional interests lie outside the sphere of IRE activities to become affiliated with certain of the IRE Professional Groups *without* first having to join the IRE itself.

This plan is aimed at those specialists in other fields of science and technology whose work touches upon our own electronics and communications field only in specialized areas. In effect, the IRE is extending the specialized services of its Professional Groups to every field of science and engineering.

An outstanding example of where these services are needed may be found in the case of the medical and biological sciences. At the present time some 1400 IRE members enjoy the privileges of membership in the Professional Group on Medical Electronics. And yet there are hundreds, perhaps thousands, of medical doctors, biologists, and others to whom the activities of this Group would be of interest and value. Both they and the Group would benefit from their participation. To require these persons, who have no interest in radio engineering, to join the IRE in order to join the Group is unreasonable, and probably futile as well. In fact, it was largely to provide an answer to this particular problem that the Affiliate Plan was first conceived, although it pertains to other fields as well, such as Computers, *etc.*

The Affiliate Plan is admittedly an experiment. So far as is known, no other society has ever tried a similar scheme. The Board of Directors feels strongly that the benefits afforded by the plan justify the risk that some persons who should join the IRE will instead become Affiliates. To minimize this risk, the plan has been carefully worked out along the following lines:

1) Participation in the plan is at the option of each Professional Group. It is not expected that all Groups will adopt it; only those which feel it serves a need in their particular field.

2) Each Group interested in initiating the Affiliate Plan must submit to the Chairman of the Professional Groups Committee a list of accredited organizations which has been selected and approved by its Administrative Committee, for official approval by the IRE Executive Committee.

3) To be an Affiliate of a Professional Group, a person must belong to an accredited organization approved by that Group and the IRE Executive Committee. Moreover, he shall not have been an IRE member during the five years prior to his application. He may affiliate with more than one Group, provided the accredited organization to which he belongs is recognized by the Groups concerned.

4) The fee for Affiliates shall be the assessment fee of the Group, plus $4.50. The latter covers IRE subsidies to the Group, Professional Group overhead expenses borne by IRE Headquarters, and 50 cents which is to be rebated to IRE Sections for mailing and meeting costs.

5) An Affiliate will be entitled to receive the TRANSACTIONS of his Group and that part of the IRE National Convention Record pertaining to his Group. He will be eligible for a Group award, and may attend local or national meetings of the Group by payment of charges assessed Group members.

6) An Affiliate cannot serve in an elective office in the Group or Group Chapter, nor vote for candidates for these offices.

7) An Affiliate may hold an appointive office in the Group or Group Chapter.

8) An Affiliate may not receive any IRE benefits that are derived through IRE membership.

The Affiliate Plan is a bold and farsighted venture; one that recognizes and provides for the rapidly spreading influence of electronics in every walk of scientific and technological life, and one that enables the IRE to further its aims as a professional engineering society—the advancement of radio engineering and related fields of engineering and science.

On the Estimation in the Presence of Noise of the Impulse Response of a Random, Linear Filter[*]

GEORGE L. TURIN†

Summary—A sounding signal is transmitted through a trans-ssion medium which may be characterized as a linear filter ose impulse response is random, *i.e.*, drawn according to some bability law from an ensemble of possible impulse responses. the output of the medium is added random noise; the resultant veform is the received signal. A receiver is required to operate this received signal so as to make a linear, minimum-mean-iare-error estimate of the impulse response of the transmission dium.

Two problems concerned with the design of such a sounding tem are considered in this paper. The first is the determination the transfer function of the optimum linear estimating filter in receiver. The second is the optimization of the spectrum of the nsmitted sounding signal.

Introduction

NE IS OFTEN faced with the problem of making measurements on a transmission medium whose characteristics are random, or at least unknown. is problem might arise, for example, in the case of mmunication through a random-multipath medium ch as the ionosphere.[1-4] In such a case, it usually is vantageous from the point of view of system per-rmance to obtain by measurement a more exact knowl-ge of the strengths and delays of the various paths an that which is available *a priori*. Measurement ould in this case be a means to the end of more reliable mmunication. On the other hand, measurement may an end in itself, as in the radar case. Here, the trans-ssion medium consists of a group of targets (and the ace between the radar set and the targets), and one interested in measuring the delays of signals reflected om these targets.

In general, one will not be able to determine the aracteristics of the medium exactly: one must make y measurement in the presence of noise, and, barring infinite measurement time, this precludes an un-uivocal measurement. In other words, one must be ntent with *estimates* of the pertinent characteristics.

The choice of the "pertinent" characteristics and the type of estimation to be used depends on the requirements of the particular problem, as well as on mathematical tractability. In this paper, we shall assume (as we may, for instance, in the two examples cited above) that the transmission medium is linear, so that it may be completely characterized by an impulse response function. We shall consider the special problem of making a minimum-mean-square-error estimate of this impulse response by means of a linear estimating filter.

We shall, in particular, consider the measurement system of Fig. 1. A sounding signal, $x(t)$, of some finite duration, T, is assumed to be transmitted. This signal passes through the transmission medium, represented in Fig. 1 as a linear filter. In general, the impulse response of this filter may be a function of time. We shall assume here, however, that if the impulse response does vary with time, it varies slowly enough so that we may consider it essentially fixed during the transmission of the sounding signal. We may then, for our purposes, write the impulse response of the medium as a function of a single variable: $h_m(\tau)$. $h_m(\tau)$ is considered to be random, at least from the viewpoint of the transmitter and receiver; that is, it may be thought of as having been drawn, according to some probability law, from an ensemble of possible impulse responses.

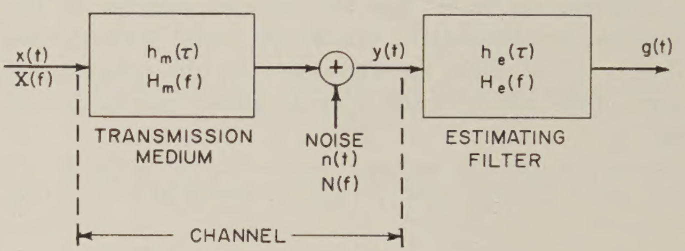


Fig. 1—The system under consideration.

After transmission through the medium, the signal is perturbed by the addition of a random noise waveform, $n(t)$, which we assume is statistically independent of $h_m(\tau)$, and is statistically stationary. The noise-perturbed signal, $y(t)$, passes into a receiver. This consists of a linear filter with impulse response, $h_e(\tau)$, whose output, $g(t)$, is an estimate (as a function of time) of the impulse response of the transmission medium. It may be noted that the receiver may be located physically at a point distant from the transmitter, as in the communication case, or at the same point as the transmitter, as in the radar case.

We shall consider two design problems connected with

[*] Manuscript received by the PGIT, July 5, 1956. The research ported in this document was supported jointly by the U. S. Army, avy, and Air Force under contract with the Mass. Inst. Tech. It abstracted from a doctoral thesis submitted to the Dept. of Elec. ng., M.I.T., May, 1956.

† Hughes Aircraft Co., Culver City, Calif. Formerly with ept. of Elec. Eng. and Lincoln Lab., M.I.T., Cambridge, Mass.

[1] G. L. Turin, "Communication Through Noisy, Random-ultipath Channels," Sc. D. thesis, M.I.T., Cambridge, Mass.; 56. Also published as Tech. Rep. 116, Lincoln Lab., M.I.T.; ay 14, 1956.

[2] G. L. Turin, "Communication through noisy, random-multi-th channels," 1956 IRE Convention Record, Part IV, pp. 4–166.

[3] R. Price, "The detection of signals perturbed by scatter and ise" IRE Trans., PGIT-4, pp. 163–170; September, 1954.

[4] W. L. Root and T. S. Pitcher, "Some remarks on statistical tection," IRE Trans., vol. IT-1, pp. 33–38; December, 1956.

the system of Fig. 1. The first is the determination of the impulse response, $h_{e_{opt}}(\tau)$, of the estimating filter which makes the mean-square difference between $g(t)$ and $h_m(t)$ a minimum for a given $x(t)$. Then we shall investigate the problem of adjusting $x(t)$, holding $h_e(\tau) = h_{e_{opt}}(\tau)$, so as to bring the mean-square error in the estimate down to an absolute minimum.

Throughout this paper we shall use Fourier transforms in which the frequency-domain variable is a cyclic, rather than a radian, frequency. That is, we shall use the transform pair

$$z(t) = \int_{-\infty}^{\infty} Z(f)e^{i2\pi ft}\,df \tag{1}$$

$$Z(f) = \int_{-\infty}^{\infty} z(t)e^{-i2\pi ft}\,dt \tag{2}$$

where $z(t)$ is the time-domain function, and $Z(f)$ its frequency-domain mate.

The Optimum Estimating Filter

Suppose we know that the impulse response of the medium lies essentially inside of some interval, say $0 \leq \tau \leq \Delta$. Then the expression for the mean-square error of the output of the estimating filter may be defined as

$$\epsilon \equiv E_{N,M}\left[\frac{1}{\Delta}\int_0^{\Delta} [g(t) - h_m(t)]^2\,dt\right] \tag{3}$$

where $E_{N,M}$ denotes a statistical average over the ensembles of possible noises and possible impulse responses of the medium. We minimize this error by varying the estimating-filter impulse-response; that is, we set

$$\delta\epsilon = 0 \tag{4}$$

where the variation ("δ") is with respect to the estimating-filter impulse-response.

It is shown in Appendix I that the transfer function of the estimating filter (*i.e.*, the Fourier transform of $h_e(\tau)$) which satisfies (4) is

$$H_{e_{opt}}(f) = \frac{1}{\Delta}\cdot\frac{X^*(f)}{\dfrac{N(f)}{|H_m(f)|^2} + \dfrac{|X(f)|^2}{\Delta}} \tag{5}$$

where the asterisk denotes "complex conjugate." In this expression, $X(f)$ is the sounding-signal voltage-density spectrum; $|H_m(f)|^2$, the average power-transmission function of the medium; and $N(f)$, the noise power-density spectrum. Thus, the only statistic of the medium which one must have *a priori* is $|H_m(f)|^2$; if we have no *a priori* knowledge of the medium, we set this equal to a constant. $|H_m(f)|^2$ may of course be a statistic derived from information gained from previous measurements.

The filter of (5) may not be physically realizable, because the condition that the impulse response of a realizable filter must be zero for negative arguments is not used in the derivation in Appendix I. However, this is not a great problem; we can usually accept some delay in obtaining our estimate of $h_m(\tau)$, and $H_{e_{opt}}(f)$ can usually be made realizable, at least to a very good approximation, by introducing sufficient delay. This delay will usually not be more than the order of T seconds.

For the no-noise case, *i.e.*, $N(f) \equiv 0$, the solution reduces to the inverse filter

$$H_{e_{opt}} = \frac{1}{X(f)}. \tag{6}$$

This is, of course, to be expected, for in this case the voltage-density spectrum at the receiver input is $H_m(f)X(f)$, and the filter of (6) restores this to just $H_m(f)$, which is the Fourier transform of $h_m(\tau)$. Thus, in the no-noise case, $h_m(\tau)$ is estimated without error.

For $N(f) \not\equiv 0$, comparison of (5) and (6) shows that the optimum filter may be expressed (see Fig. 2) as the cascade of the inverse filter of (6) and a filter with transfer function

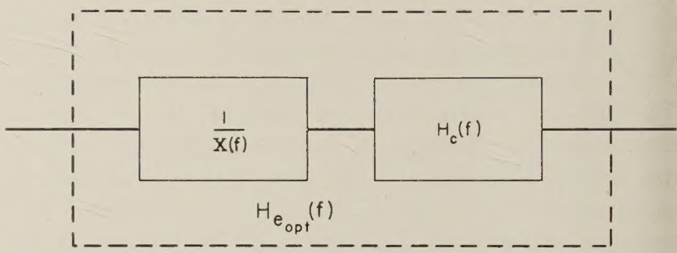


Fig. 2—Analysis of $He_{opt}(f)$ into component filters.

$$H_c(f) = \frac{\dfrac{|X(f)|^2}{\Delta}}{N_r(f) + \dfrac{|X(f)|^2}{\Delta}} \tag{7}$$

where we have set

$$N_r(f) = \frac{N(f)}{|H_m(f)|^2}.$$

($N_r(f)$ is thus the noise power-density spectrum, referred back to the transmitter through the *average* medium.) The output of the inverse filter in Fig. 2 contains $h_m(\tau)$ but it also contains a large amount of noise, especially at those frequencies where $X(f)$ is small. The second filter attenuates the noise, but in doing so, smears $h_m(\tau)$. The optimization procedure may be thought of as one which makes the best compromise between eliminating noise and keeping $h_m(\tau)$ undistorted. $H_c(f)$ is a zero-phase filter, as one would expect, for any other phase (except a linear one, which is a trivial exception) would distort the desired output, $h_m(\tau)$, without helping to attenuat e effect of the noise, which has random phase anyway. If Δ is of the order of magnitude of T, then the second rm in the denominator of (5) is roughly the transmitted wer-density spectrum. If this is small, for all f, com- red to $N_r(f)$ (or, equivalently, if the average received- gnal power-density spectrum is small compared to (f)), then (5) becomes

$$H_{e_{opt}}(f) = \frac{1}{\Delta} \frac{X^*(f)}{N_r(f)}. \qquad (8)$$

$N_r(f)$ is constant with frequency, which will occur, for ample, if the noise is white and we have no *a priori* owledge of the medium, then the filter of equation) is "matched"[5] to the sounding signal; that is, $H_{e_{opt}}(f)$ proportional to the complex conjugate of $X(f)$, or, uivalently, $h_{e_{opt}}(\tau)$ is proportional to $x(-\tau)$, the unding signal reversed in time.

As has been suggested in the Introduction, the above sults may have application to multipath-communication d radar problems. In both of these cases, the impulse sponse of the medium may be represented to a good proximation as a sequence of impulses (Dirac delta nctions), as in Fig. 3(a). This figure is drawn for a four- th (four-target) medium. The signal component, $g_s(t)$, the estimate (or, more practically, the *envelope* of the gnal component) will look something like Fig. 3(b); a ise component will generally be superimposed on $g_s(t)$. he nonzero widths of the pulses in Fig. 3(b) result from e fact that the bandwidths of the sounding signal and the estimating filter are usually finite; in fact, the pulse dths will generally be of the order of the reciprocal of e sounding-signal bandwidth.] The strengths, a_i, and e delays, τ_i, of the various paths (targets) may be timated from the $g(t)$ waveform (although only the latter rameters are of interest in the radar case). Further ne) delay information may be obtained from an exami- tion of the phase of $g(t)$ at the envelope peaks.

In the multipath or radar context, in the one-path or rget case, the filter derived in this paper bears a relation- ip to other filters derived by Van Vleck and Middleton[5] d Dwork.[6] These latter filters may be thought of as sed on a criterion which requires that at some time rresponding to the path (target) delay, the filter output as accurate an estimate as possible of the path (target) rength, without regard to the values of the output at her times. The criterion adopted in this paper requires rther that the filter output be simultaneously as close zero as possible at all other times. Eq. (8), for the one- th case $[N_r(f) = N(f)]$, indicates that the two criteria ad to the same filter only in the limit of small signal-to- ise ratio.

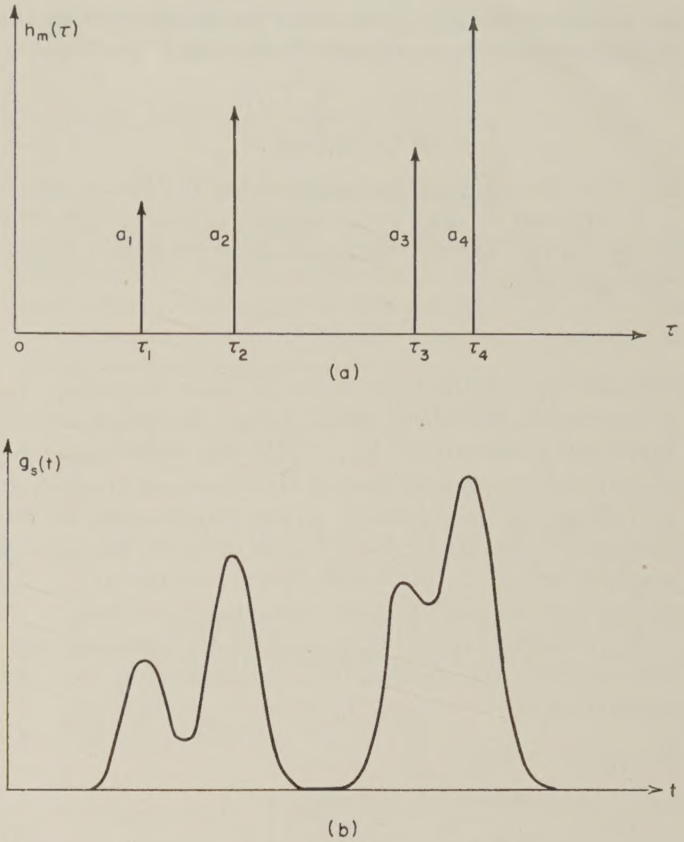


Fig. 3—(a) A typical multipath (multi-target) impulse response. (b) The signal component of an optimum estimate of (a).

The Optimum Sounding-Signal Spectrum

So far we have considered $X(f)$ to be arbitrary. Now let us, while keeping $H_e(f) = H_{e_{opt}}(f)$, solve for the $X(f)$ which minimizes ϵ, subject to the constraints that the energy in $x(t)$ be fixed and that $X(f)$ lie within a given band. That is, let us set

$$\int_{-\infty}^{\infty} | X(f) |^2 \, df = K \qquad (9)$$

and

$$X(f) \equiv 0 \qquad \text{for } f \text{ not in } F_1 \qquad (10)$$

where F_1 is the permitted band of frequencies, and solve the equation

$$\delta(\epsilon_m + \lambda K) = 0 \qquad (11)$$

where λ is some constant.[7] ϵ_m is the minimum mean-square error for an arbitrary $X(f)$.

The solution to (11) is shown in Appendix II to be

$$X_{opt}(f) = \sqrt{\Delta} e^{-i\beta(f)} \cdot \begin{cases} [N(f)]^{1/4} \left[\dfrac{1}{\Delta\sqrt{\lambda}} - \dfrac{\sqrt{N(f)}}{| H_m(f) |^2} \right]^{1/2} & f \text{ in } F_2 \\ 0 & f \text{ in } F_2'. \end{cases} \qquad (12)$$

[5] J. H. Van Vleck and D. Middleton, "A theoretical comparison the visual, aural, and meter reception of pulsed signals in the esence of noise," *J. Appl. Phys.*, vol. 17, pp. 940–971; November, 46.

[6] B. M. Dwork, "Detection of a pulse superimposed on fluctua- n noise," Proc. IRE, vol. 38, pp. 771–774; July, 1950.

[7] F. B. Hildebrand, "Methods of Applied Mathematics," Prentice-Hall, Inc., New York, N.Y.; 1952. See sec. 2.6.

In this equation, $\beta(f)$ is an arbitrary phase function, F_2 is the set of all frequencies in F_1 for which

$$\frac{1}{\Delta\sqrt{\lambda}} \geq \frac{\sqrt{N(f)}}{|\,H_m(f)\,|^2},$$

and F_2' is the set of all frequencies not in F_2. The constant, λ, is adjusted to satisfy the energy constraint, (9). The interpretation of (12) confirms one's intuitive notions. The first factor,

$$[N(f)]^{1/4},$$

indicates that if the noise power at some frequency, f_1, is very small, then little signal energy is needed at that frequency to determine $H_m(f_1)$. On the other hand, the second factor indicates that if the noise power at f_1 is very large, or the average power transmission of the medium is very small, then it is a waste of the limited available energy to put much, if any, energy at f_1; the energy may be used to more advantage elsewhere.

If we place (12) in (5), we obtain the optimum estimating-filter transfer-function corresponding to the optimum $X(f)$:

$$H_{e_{\text{opt}}}(f) = \sqrt{\lambda\Delta}e^{+i\beta(f)}$$
$$\cdot\begin{cases} [N(f)]^{-1/4}\left[\dfrac{1}{\Delta\sqrt{\lambda}} - \dfrac{\sqrt{N(f)}}{|\,H_m(f)\,|^2}\right]^{1/2} & f \text{ in } F_2 \\ 0 & f \text{ in } F_2'. \end{cases} \quad (13)$$

We have assumed here that $N(f)$ is nonzero at all frequencies. The optimum filter of (13), as we should expect, has large gain at frequencies with small noise, and small, or zero, gain at frequencies with large noise.

The mean-square error corresponding to (12) and (13) is, from Appendix II:

$$\epsilon_{mm} = \left[\int_{F_2} \sqrt{N(f)\lambda}\, df + \frac{1}{\Delta}\int_{F_2'} \overline{|\,H_m(f)\,|^2}\, df\right]. \quad (14)$$

The first term in this expression is the contribution to the error of the noise and smear components of the estimate which arise within the pass band of $H_{e_{\text{opt}}}(f)$. The second is the smear contribution arising from the complete lack of an estimate of $H_m(f)$ in the stopband of $H_{e_{\text{opt}}}(f)$.

We may actually be interested only in estimating $H_m(f)$ within the transmission band, F_1, instead of for all f as we have done. That is, we may be interested in estimating the instantaneous impulse response of an equivalent medium which has zero transmission outside of F_1. This imposes no additional problem, however; it is easy to see that the result of (13) is optimum in this case also, since it is independent of values of $H_m(f)$ outside of the transmission band. The error of (14) also obtains in this case, except that the second integral is now taken over only those frequencies in F_2' which are within the transmission band (*i.e.*, the intersection of F_1 and F_2').

We note as a final point that if the noise is white, *i.e.* $N(f)$ is constant, at least over the transmission band, the optimum estimator of (13) is proportional to the complex conjugate of the sounding-signal spectrum of (12). That is, in the white-noise case the optimum estimator is matched to the optimum sounding signal, regardless of the form of the average power-transmission spectrum of the medium, $\overline{|\,H_m(f)\,|^2}$.

Appendix I

Derivation of $H_{e_{\text{opt}}}(f)$

We desire the solution of (4):

$$\delta\epsilon = \delta E_{N,M}\left[\frac{1}{\Delta}\int_0^{\Delta} [g(t) - h_m(t)]^2\, dt\right] = 0. \quad (15)$$

Expanding this, we obtain

$$E_{N,M}\left[\int_0^{\Delta} g(t)\, \delta g(t)\, dt\right] = E_{N,M}\left[\int_0^{\Delta} h_m(t)\, \delta g(t)\, dt\right] \quad (16)$$

since $\delta h_m(t) = 0$. Now $g(t)$, as indicated in Fig. 1, is the output of the estimating filter, whose unit-impulse response is $h_e(\tau)$. Since $y(t)$ is the filter input,

$$g(t) = \int_{-\infty}^{\infty} h_e(\tau)y(t - \tau)\, d\tau. \quad (17)$$

$y(t)$ is, in turn, the sum of the noise, $n(t)$, and the output of the transmission medium, whose unit-impulse response, $h_m(\tau)$, is to be estimated. That is

$$y(t) = \int_{-\infty}^{\infty} h_m(\tau)x(t - \tau)\, d\tau + n(t) \quad (18)$$

where $x(t)$ is the channel input (sounding signal). Combining (17) and (18):

$$g(t) = \iint_{-\infty}^{\infty} h_e(\tau)h_m(\sigma)x(t - \tau - \sigma)\, d\sigma\, d\tau$$
$$+ \int_{-\infty}^{\infty} h_e(\tau)n(t - \tau)\, d\tau. \quad (19)$$

The variation of $g(t)$ is thus

$$\delta g(t) = \iint_{-\infty}^{\infty} h_m(\sigma)x(t - \tau - \sigma)\, \delta h_e(\tau)\, d\sigma\, d\tau$$
$$+ \int_{-\infty}^{\infty} n(t - \tau)\, \delta h_e(\tau)\, d\tau. \quad (20)$$

Remembering that $n(t)$ and $h_m(\tau)$ are independent, and assuming that $E_N[n(t)] = 0$, we obtain for the right-hand side of (16)

$$E_M\left[\int_0^{\Delta} \iint_{-\infty}^{\infty} h_m(t)h_m(\sigma)x(t - \tau - \sigma)\, \delta h_e(\tau)\, d\sigma\, d\tau\, dt\right]. \quad (21)$$

Similarly, the left-hand side of (16) is

$$E_M\Bigg[\int_0^\Delta \iiiint_{-\infty}^{\infty} h_e(\tau')h_m(\sigma')h_m(\sigma)x(t-\tau'-\sigma')$$

$$\cdot x(t-\tau-\sigma)\ \delta h_e(\tau)\ d\sigma'\ d\tau'\ d\sigma\ d\tau\ dt\Bigg]$$

$$+ E_M\Bigg[\Delta \iint_{-\infty}^{\infty} h_e(\tau')\phi_n(\tau-\tau')\ \delta h_e(\tau)\ d\tau'\ d\tau\Bigg]. \qquad (22)$$

In deriving (22) we have assumed statistically-stationary noise with autocorrelation function

$$\phi_N(\tau) = E_N[n(t)n(t+\tau)]. \qquad (23)$$

Since we have assumed that Δ is essentially greater than the duration of $h_m(\tau)$, the limits on the first integral sign in (21) may be extended to $(-\infty, +\infty)$ without changing the value of the integral. The same statement may be made approximately about the first term in (22); for this term is the integral of the product of the "signal" component[8] of the estimate of $h_m(\tau)$ and the variation of this component, and one would not expect the signal component to last appreciably longer than $h_m(\tau)$ itself.[9]

Thus extending the limits, and equating (21) and (22), we get

$$\int_{-\infty}^{\infty} \delta h_e(\tau)\ d\tau\ E_M\Bigg[\iiiint_{-\infty}^{\infty} h_e(\tau')h_m(\sigma')h_m(\sigma)x(t-\tau'-\sigma')$$

$$\cdot x(t-\tau-\sigma)\ d\sigma'\ d\tau'\ d\sigma\ dt + \Delta\int_{-\infty}^{\infty} h_e(\tau')\phi_N(\tau-\tau')\ d\tau'$$

$$- \iint_{-\infty}^{\infty} h_m(t)h_m(\sigma)x(t-\tau-\sigma)\ d\sigma\ dt\Bigg] = 0. \qquad (24)$$

We shall neglect the physical realizability condition, $h_e(\tau) = 0$ for $\tau < 0$. Then $\delta h_e(\tau)$ is arbitrary for all τ, and in order for (24) to be satisfied, the factor which multiplies $\delta h_e(\tau)$ must be zero for all τ. Setting it equal to zero, Fourier transforming the resulting equation, and averaging over the ensemble of all possible transmission-medium impulse responses, we obtain

$$H_{e\,\mathrm{opt}}(f)\ \overline{|\,H_m(f)\,|^2}\ |\,X(f)\,|^2 + \Delta H_{e\,\mathrm{opt}}(f)N(f)$$

$$- \overline{|\,H_m(f)\,|^2}\ X^*(f) = 0. \qquad (25)$$

Eq. (5) follows from this immediately. In order to show that this solution yields a minimum, rather than a maximum or inflectional, error, one merely finds the second variation of ϵ, and shows that this is positive for $H_e(f) = H_{e\,\mathrm{opt}}(f)$.

Appendix II

Derivation of $X_{\mathrm{opt}}(f)$

To derive (12) for the optimum spectrum of $x(t)$, we start with (3) for the mean-square error. Using (19) in this, and remembering that the average of the bracketed expression in (24) is identically zero, we obtain for the minimum mean-square error

$$\epsilon_m = \frac{1}{\Delta} E_M\Bigg[\int_{-\infty}^{\infty} h_m^2(t)\ dt$$

$$- \iiint_{-\infty}^{\infty} h_m(t)h_m(\sigma)h_{e\,\mathrm{opt}}(\tau)x(t-\tau-\sigma)\ d\sigma\ d\tau\ dt\Bigg]. \qquad (26)$$

ϵ_m is also expressible in terms of frequency-domain functions; using Parseval's theorem in (26) and averaging:

$$\epsilon_m = \frac{1}{\Delta}\Bigg[\int_{-\infty}^{\infty} \overline{|\,H_m(f)\,|^2}\ df$$

$$- \int_{-\infty}^{\infty} H_{e\,\mathrm{opt}}(f)\ \overline{|\,H_m(f)\,|^2}\ X(f)\ df\Bigg]. \qquad (27)$$

Using (5) for $H_{e\,\mathrm{opt}}(f)$, (27) becomes

$$\epsilon_m = \frac{1}{\Delta}\int_{-\infty}^{\infty} \overline{|\,H_m(f)\,|^2}\left[1 - \frac{|\,X(f)\,|^2}{N_r(f)\Delta + |\,X(f)\,|^2}\right] df \qquad (28)$$

where

$$N_r(f) = \frac{N(f)}{\overline{|\,H_m(f)\,|^2}}.$$

We now constrain the energy in the transmitted waveform to be constant:

$$\int_{-\infty}^{\infty} |\,X(f)\,|^2\ df = K. \qquad (29)$$

In order to find the optimum $X(f)$, we must solve the variational problem[7]

$$\delta(\epsilon_m + \lambda K) = 0 \qquad (30)$$

where λ is some constant. Using (28) and (29), (30) becomes

$$\int_{-\infty}^{\infty}\left[\lambda - \frac{N_r(f)\ \overline{|\,H_m(f)\,|^2}}{[N_r(f)\cdot\Delta + |\,X(f)\,|^2]^2}\right]\delta\,|\,X(f)\,|^2\ df = 0. \qquad (31)$$

If we constrain $|\,X(f)\,|^2$ to be zero outside a certain band, F_1 [*cf.* (10)], then $\delta\,|\,X(f)\,|^2$ is also zero there, and (31) is satisfied for those frequencies. For frequencies within the band, on the other hand, we must try to set the bracketed

[8] *i. e.*, the first term in (19).

[9] The approximation here is very good if the reciprocal of the bandwidth of $x(t)$ is small compared to the duration of $h_m(\tau)$. This condition is also necessary if the fine detail in $h_m(\tau)$ is to be highly resolved by the sounding system.]

term in the integrand equal to zero.[10] This leads to the equation

$$\lambda \mid X(f) \mid^4 + 2\lambda\Delta N_r(f) \mid X(f) \mid^2 + N_r(f)[\lambda\Delta^2 N_r(f) - \overline{\mid H_m(f) \mid^2}] = 0. \qquad (32)$$

If (32) and (29) can be simultaneously satisfied within the band F_1 by a non-negative function $\mid X(f) \mid^2$, then the solution is complete. If, however, there are frequencies at which $\mid X(f) \mid^2$ would be negative, then it may be shown[1] that the correct solution is to satisfy (32) and (29) simultaneously at all frequencies for which $\mid X(f) \mid^2$ turns out non-negative, while at the same time setting $\mid X(f) \mid^2$ equal to zero at all other frequencies; this is the solution indicated in (12).

Finally, (14) follows directly on substitution of (12) into (28).

Acknowledgment

The author would like to extend his thanks to Prof. R. M. Fano and Dr. W. B. Davenport, Jr. of M.I.T. for their help and guidance in this research.

[10] One might be tempted to obtain another solution, $X(f) = 0$, by writing $\delta \mid X(f)\mid^2 = 2 \mid X(f)\mid \delta \mid X(f)\mid$. This is a spurious solution, however, for the problem is actually phrased completely in terms of $\mid X(f)\mid^2$ [*cf.* (28) and (29)]. The second solution would disappear if we replaced $\mid X(f)\mid^2$ by, say, $S(f)$.

The Output Signal-to-Noise Ratio of Correlation Detectors*

PAUL E. GREEN, JR.†

Summary—**Expressions are derived for the output signal-to-noise ratio of a correlation detector when the two input functions to be correlated differ only by the presence of an arbitrary linear filter in each path, and the addition of noise to each. It is assumed that the signal and noises are Gaussian with arbitrary power density spectra, and the integration is performed by a filter of arbitrary transfer function. Two types of correlation detectors are distinguished, the low-pass detector in which the integrator is a low-pass filter, and the band-pass detector in which one of the two input functions is deliberately displaced in frequency by Δ and the integrator is therefore a band-pass filter tuned to Δ. Output signal-to-noise ratio expressions for the two types are almost identical.**

Introduction

The use of cross correlation for the detection of signals in noise has been treated from several points of view. Lee, Cheatham, and Wiesner[1] discussed the use of a sampling correlator for this function, and Fano[2] and Davenport[3] extended the treatment to the type of correlator which compared continuous waveforms rather than sample values of these waveforms.[4] The present paper treats the continuous correlator in a more general fashion than was done in the papers of Fano and Davenport.

The action of a correlation detector is to multiply two waveforms together and perform an integration or smoothing of the product. The situation is depicted in Fig. 1. The detector consists of the multiplier and integrating filter. Usually there is some relationship between the two input waveforms [indicated in Fig. 1 as $u_1(t)$ and $u_2(t)$] causing them to have a nonzero correlation. As a matter of fact, in most practical cases, they are the same waveform $x(t)$ disturbed by the addition of noise [$n_1(t)$ and $n_2(t)$], and possibly also distorted dissimilarly in some other way. The effect of the noise at the detector

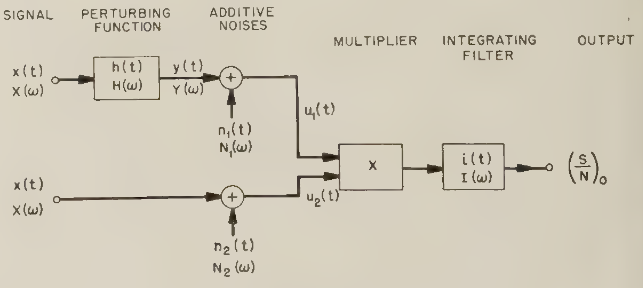


Fig. 1—Correlation detector.

* Manuscript received by the PGIT, January 17, 1957. The research in this document was supported jointly by the U. S. Army, Navy, and Air Force under contract with Mass. Inst. of Tech.

† Lincoln Lab., M.I.T., Lexington, Mass.

[1] Y. W. Lee, T. P. Cheatham, and J. B. Wiesner, "The Application of Correlation Functions in the Detection of Small Signals in Noise," Tech. Rep. No. 141, Res. Lab. of Electronics, M.I.T.; October 13, 1949.

[2] R. M. Fano, "Signal to Noise Ratio in Correlation Detectors," Tech. Rep. No. 186, Res. Lab. of Electronics, M.I.T.; February 19, 1951.

[3] W. B. Davenport, Jr., "Correlator Errors Due to Finite Observation Intervals," Tech. Rep. No. 191, Res. Lab. of Electronics, M.I.T.; March 8, 1951.

[4] The two are, of course, equivalent if the sampling rate is sufficiently high.

tput tends to zero as the duration of the integration ows indefinitely. In a particular system application, e practical necessity of limiting this integration time to ite values produces a finite signal-to-noise ratio at the tput. It is this signal-to-noise ratio that specifies the rformance of the detector.

This paper presents an analysis of the output signal--noise ratio of a correlation detector under the following sumptions:

1) The signal $x(t)$, and the noises $n_1(t)$ and $n_2(t)$, which are inserted additively, are independent stationary and ergodic random functions of time with Gaussian first- and second-order amplitude distributions, and with Fourier-transformable power density spectra $X(\omega)$, $N_1(\omega)$, and $N_2(\omega)$, respectively, all confined to $2\pi W$, a closed interval in ω. For simplicity of computation, a single-sided frequency spectrum is used for all functions of time, whereas a double-sided representation is employed for network system functions.
2) The only perturbing function causing any dissimilarity in $u_1(t)$ and $u_2(t)$ not accounted for by the added noises, is a linear, but not necessarily physically realizable, time-invariant network with a complex system function $H(\omega) = |H(\omega)| e^{i\eta(\omega)t}$, Fourier-transformable into the impulse response $h(t)$.[5]

 this treatment, such a filter is included in only one of e two correlator inputs, since the presence of a different ter in each input can be treated by readjusting the signal t) to represent one correlator input while lumping the fference in the two perturbing filters into the single ter shown in Fig. 1. The presence in the analysis of the rturbing filter can be exploited in many ways. For ample, we will use it later to represent a time shift etween the two signal components at the correlator puts. It can also be used to represent the presence of ssimilar filtering of the two signals, and of such propagaion effects as time invariant multipath and dispersion. hese latter details will not be pursued in this paper.

3) The multiplier is an ideal four-quadrant multiplier whose output is the instantaneous product of the values of u_1 and u_2.
4) The integrating filter is a realizable two-terminal pair device with complex system function $I(\omega)$, Fourier-transformable into the filter impulse response $i(t)$.

 the particular case where $i(t)$ is a rectangular pulse of uration T, we will call the filter an *ideal* integrator of tegration time T. For other appropriate forms of $i(t)$, there will turn out to be an *effective* integration time T equal to the reciprocal of the effective noise bandwidth of the filter.

Our problem will be to compute the output signal-to-noise ratio, that is, the ratio of the square of the dc output voltage from the integrating filter to the fluctuation power at the same point. Specifically, we proceed as follows: An infinite ensemble of sets of the three functions $x(t)$, $n_1(t)$, and $n_2(t)$ is considered, all of period θ, where θ is long compared to the duration of significant values of the filter response $i(t)$. For one set of the three functions x, n_1, and n_2 out of the ensemble of such sets, each function is expanded in a Fourier series, thus giving three line spectra, where each line represents a Fourier amplitude and phase coefficient pair. Then the operations indicated in Fig. 1 are carried out on these functions to derive the amplitude and phase line spectrum at the *multiplier* output. The power in each spectral line is then found from the square of the amplitude, and the ensemble average power in the various spectral lines is then computed based on known statistical properties of the Fourier coefficients. Then $|I(\omega)|^2$, the squared magnitude of the integrator's system function, is applied to this discrete power spectrum at the multiplier output. The final step is to allow the Fourier period θ to grow without limit, whereupon the summations involving the Fourier coefficients become integrals involving the power density spectra $X(\omega)$, $N_1(\omega)$, and $N_2(\omega)$. The power density spectrum of the integrator output is a line impulse at the origin representing the dc output signal, plus a continuous spectrum representing fluctuations. The ratio of these two is the desired output signal-to-noise ratio.

In the next section we will treat in this way the correlation detector just as shown in Fig. 1. Then we will treat a variation of this in which the two input functions are displaced in frequency and the integrating filter is a band-pass rather than a low-pass filter. We distinguish between these two types by calling them *low-pass* and *band-pass* type correlation detectors, respectively.

Low-Pass Correlation Detector

For each set of the ensemble of signal and noise functions, we have the following Fourier expansions

$$x(t) = \sum_{i=1}^{\theta W} \xi_i \cos(\omega_i t + \phi_i), \tag{1}$$

$$n_1(t) = \sum_{i=1}^{\theta W} \nu_i \cos(\omega_i t + \gamma_i), \tag{2}$$

and

$$n_2(t) = \sum_{i \cdot 1}^{\theta W} \mu_i \cos(\omega_i t + \delta_i). \tag{3}$$

Then

$$u_1(t) = y(t) + n_1(t)$$

$$= \sum_{i=1}^{\theta W} \xi_i h_i \cos(\omega_i t + \phi_i + \eta_i) + n_1(t) \tag{4}$$

[5] The presence of a perturbing function in the form of a time variant nonlinear distortion of amplitudes has been treated in e following references: J. J. Bussgang, "Crosscorrelation Functions Amplitude-Distorted Gaussian Signals," Tech. Rep. No. 216, es. Lab. of Electronics, M.I.T.; March 26, 1952, and R. D. Luce, Amplitude distorted signals," *R. L. E. Quarterly Prog. Rep.*, pp. 7–41; April 15, 1953.

and

$$u_2(t) = x(t) + n_2(t) \tag{5}$$

where ω_0 represents the lower edge of the band of width W cps placed to include all signal and noise components, $\omega_i = \omega_0 + 2\pi i/\theta$, and h_i and η_i are the perturbing filter amplitude and phase functions $| H(\omega_i) |$ and arc $[H(\omega_i)]$, respectively. It is known[6] that when $x(t)$ is a stationary Gaussian random process having power density spectrum $X(\omega)$, each ξ_i has a Rayleigh distribution over the ensemble with

$$\overline{\xi_i^2} = 2X(\omega_i)\Delta\omega \tag{6}$$

and

$$\overline{\xi_i^4} = 2\overline{(\xi_i^2)}^2 = 8X^2(\omega_i)(\Delta\omega)^2 \tag{7}$$

for sufficiently small $\Delta\omega = 2\pi/\theta$. If $i \neq j$, ξ_i and ξ_j are independent, as are ϕ_i and ϕ_j. The phase angle ϕ_i has a probability distribution which is flat from $-\pi$ to $+\pi$. Similar statements hold for $n_1(t)$ in relation to ν_i, γ_i, and the power density spectrum $N_1(\omega)$ and for $n_2(t)$ in relation to u_i, δ_i, and $N_2(\omega)$.

At the multiplier output there will be four distinct contributions from the product

$$u_1(t)u_2(t) = [y(t) + n_1(t)][x(t) + n_2(t)], \tag{8}$$

which we will designate the $X \times Y$, $X \times N_1$, $Y \times N_2$, and $N_1 \times N_2$ terms, using the subscripts I, II, III, and IV, respectively. Fig. 2 depicts the line spectrum of the power of one ensemble member of $x(t)$, $n_1(t)$, and $n_2(t)$, and shows how $u_1(t) = y(t) + n_1(t)$ and $u_2(t) = x(t) + n_2(t)$. It also shows how the operation of the multiplier produces the signal and noise components. (There are also components about the double frequency that will be ignored since they can be presumed to lie outside the integrating filter pass band.) Note the large dc output signal resulting from the coherence between $x(t)$ and $y(t)$.

Appendix I contains the detailed bookkeeping involved in carrying out the steps outlined in the introduction. For a given ensemble member, the discrete Fourier amplitude and phase spectra at the multiplier output are first computed [see (23), (29), and (34)]. From these, the corresponding discrete ensemble average power spectra are deduced, making use of the statistical independence between the signals and noises $x(t)$, $n_1(t)$, and $n_2(t)$. The lack of such independence in $x(t)$ and $y(t)$ leads to a behavior in the $X \times Y$ term [(25) and (28)] that is entirely different from that in any of the other three [(31) through (33) and (35) through (37)]. In particular, the $X \times Y$ term contains a component at dc [second term of (25)] that does not approach zero in the limit as $\theta \to \infty$. This is the correlated signal output. By multiplying this by the integrating filter's dc response, and then letting $\theta \to \infty$, we have the final signal output power, S_0 (40). By multiplying all the other terms by the correct value of the integrating filter response and then taking the limit, we have the noise output power, N_0 (41). The ratio of the two is the desired signal-to-noise ratio result.

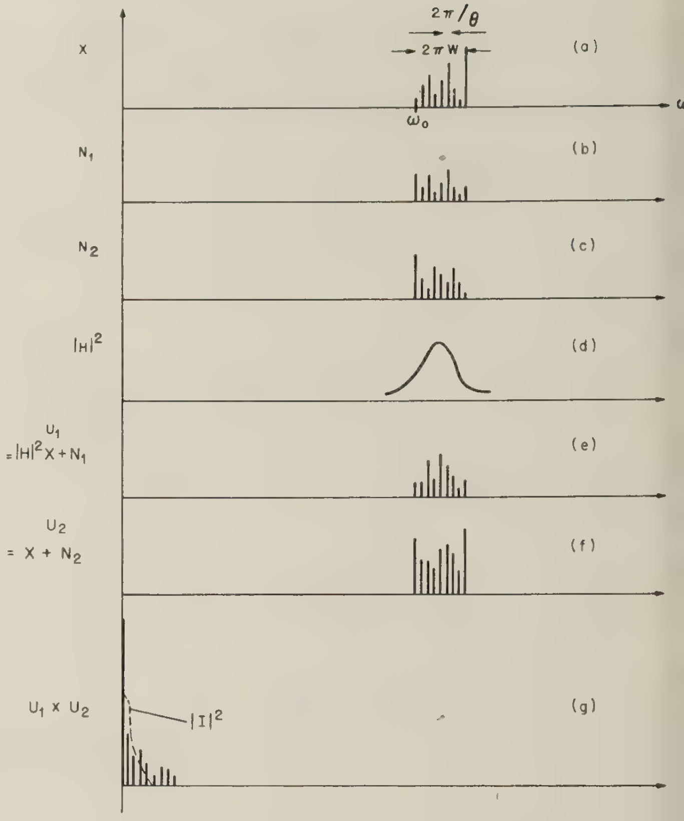


Fig. 2—Showing the expansion of $x(t)$, $n_1(t)$, and $n_2(t)$ of finite period θ into Fourier series; how the multiplication operation produces a strong dc spike (the output signal) from the coherence between $x(t)$ and $y(t)$; and how the integrating filter passes this spike, plus a small portion of the other components (the output fluctuations).

$$\left(\frac{S}{N}\right)_0 = 2\left\{| I(0) | \int_0^\infty X(\omega)Re[H(\omega)]\,d\omega\right\}^2 \cdot \left\{\int_0^\infty | I(\omega) |^2 [Re([X(\omega)H(\omega)]*[X(\omega)H(\omega)]) + X(\omega)*N_1(\omega) + (X(\omega) \mid H(\omega) \mid^2)*N_2(\omega) + N_1(\omega)*N_2(\omega)]\,d\omega\right\}^{-1} \tag{9}$$

in which the operation "*" is defined by

$$A(\omega)*B(\omega) = \int_0^\infty [A(u)B(u+\omega) + A(u+\omega)B(\omega)]\,du \tag{10}$$

and the abbreviation Re means "real part of."

Band-Pass Correlation Detector

Suppose in Fig. 1 we use as the lower detector input signal $u_2(t)$, exactly the same function as before, except displaced downward in radian frequency by Δ. This can

[6] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, sec. 2.8.; July, 1944 and January, 1945.

be taken into account by rewriting the Fourier expansions (3) and (5) as

$$n_2(t) = \sum_{i=1}^{\theta W} \mu_i \cos(\omega_i t - \Delta t + \delta_i) \tag{3a}$$

and

$$u_2(t) = \sum_{i=1}^{\theta W} \xi_i \cos(\omega_i t - \Delta t + \phi_i) + n_2(t). \tag{5a}$$

We assume that the integrating filter is tuned to Δ (that is, the frequency of maximum response of the filter is Δ) and that Δ is at least equal to $2\pi W$ the bandwidth of significant values of signal and noises. We now seek as $(S/N)_0$ the ratio of one half the squared amplitude of the difference frequency tone at Δ to the remaining power, representing fluctuations.

Notice that the band-pass detector does not perform the true operation of cross-correlation. The cross-correlation function of two signals with nonoverlapping spectra is identically zero. Nevertheless, we shall see that the band-pass scheme, as a detector of signals is in most respects equivalent to the correlation operation.

In Appendix II is computed the power in the various lines at the multiplier output which is examined now in the neighborhood of Δ. The desired output signal-to-noise ratio is

$$\left(\frac{S}{N}\right)_0 = \left\{\left[\int_0^\infty X(\omega)Re[H(\omega)]\,d\omega\right]^2 + \left[\int_0^\infty X(\omega)Im[H(\omega)]\,d\omega\right]^2\right\} \cdot \left\{\int_0^\infty |\, I(\omega)\,|^2 \int_0^\infty \Big[X(u)X(u+\omega)\,|\, H(u+\omega)\,|^2 + X(u)N_1(u+\omega) + X(u)\,|\, H(u)\,|^2\, N_2(u+\Delta+\omega) + N_1(u)N_2(u+\Delta+\omega)\Big]\,du\,d\omega\right\}^{-1}. \tag{11}$$

Self-Noise

Attention should be drawn to the first denominator integrand terms in both (9) and (11), the ones involving only X, $|\, H\,|$, and η. These terms signify the presence of fluctuation components in the output, even for both $N_1(\omega)$ and $N_2(\omega)$ identically zero. This *self-noise* component is due to the fact that the signal output of the detector is a finite-time measurement of a property of a random function, (in this case it is a short-time cross-correlation[7]). Therefore, the measurement itself fluctuates (although less and less with increased integrating time), and this fluctuation is the self-noise. The difference in the self-noise component for the two types of detectors is due to the fact that in the band-pass detector, spectral components symmetrically placed about the difference frequency Δ add on a power basis, whereas in the low-pass detector, they are really at the same frequency, and thus add on a voltage basis. The other three denominator integrand terms are seen to be composed of the convolutions of the corresponding power density spectra.

Particular Cases

The $(S/N)_0$ equations (9) and (11) are applicable to any situation obeying the original assumptions. In order to see what these equations mean in physical terms, let us first make the further assumption that the integrating filter bandwidth is small compared to the signal and noise spectra, which in turn are reasonably continuous functions of ω within their bandwidth $2\pi W$. This will insure that the output noise spectra are approximately constant across the integrator pass band. Then (9) and (11) for low- and band-pass detectors can be rewritten

$$\left(\frac{S}{N}\right)_0 = \frac{\left\{\int_0^\infty X(\omega)Re[H(\omega)]\,d\omega\right\}^2}{\int_0^\infty [X^2(\omega)Re^2[H(\omega)] + X(\omega)N_1(\omega) + X(\omega)\,|\, H(\omega)\,|^2\, N_2(\omega) + N_1(\omega)N_2(\omega)]\,d\omega}\,\frac{1}{W_f} \tag{12}$$

and

$$\left(\frac{S}{N}\right)_0 = \frac{\left\{\int_0^\infty X(\omega)Re[H(\omega)]\,d\omega\right\}^2 + \left\{\int_0^\infty X(\omega)Im[H(\omega)]\,d\omega\right\}^2}{\int_0^\infty [X^2(\omega)\,|\, H(\omega)\,|^2 + X(\omega)N_1(\omega) + X(\omega)\,|\, H(\omega)\,|^2\, N_2(\omega) + N_1(\omega)N_2(\omega)]\,d\omega}\,\frac{1}{W_f} \tag{13}$$

respectively. In these two equations the effect of the integrating filter is given by the single quantity

$$W_f = |\, I_{\max}(\omega)\,|^{-2} \int_0^\infty |\, I(\omega)\,|^2\, d\omega \tag{14}$$

which will be recognized as the *effective noise bandwidth*[8] of the filter in radians per second. The reciprocal of $W_f/2\pi$, the effective noise bandwidth (in cycles per second), can be called the *effective integration time*[9] of the filter since it produces the same $(S/N)_0$ as an ideal integrating filter having a rectangular impulse response of duration $2\pi/W_f$.

[7] R. M. Fano, "Short-time autocorrelation functions and power spectra," *J. Acoust. Soc. Am.*, vol. 22, pp. 546–50; September, 1950.

[8] J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," M.I.T. Radiation Lab. Series, vol. 24, McGraw-Hill Book Co., Inc., New York, N.Y., p. 176; 1954.

[9] To the knowledge of the author, the equivalence between reciprocal of noise bandwidth and the effective integrating time of a filter (the time duration of an ideal filter having the same effect) was first derived by B. L. Basore (private communication).

One can verify this by substituting the Fourier transform of such an ideal impulse response into (14). Ideal impulse responses of low-pass and band-pass filters are depicted in Fig. 3.

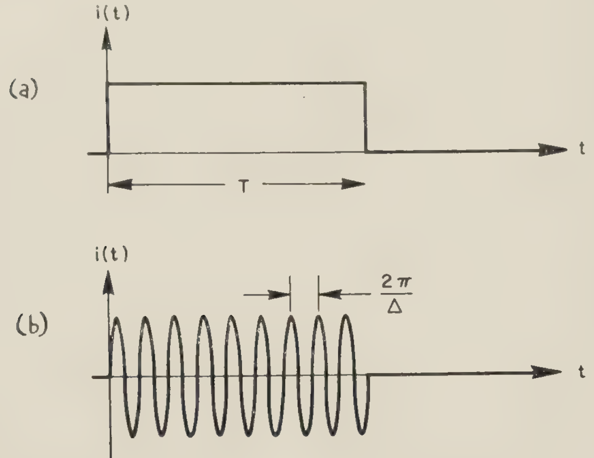


Fig. 3—Impulses response of ideal integrators (a) low-pass, (b) band-pass (the phase angle and number of cycles is immaterial for present purposes).

The $(S/N)_0$ equations (12) and (13) can be applied to a number of even more restrictive cases that are also interesting. As an example, the case where $H(\omega) \equiv$ unity and the three time functions (x, n_1, and n_2) have the same spectral shapes gives

$$\left(\frac{S}{N}\right)_0 = \frac{W_x}{FW_f}\left[\epsilon + \frac{1}{\rho_1} + \frac{1}{\rho_2} + \frac{1}{\rho_1\rho_2}\right]^{-1} \tag{15}$$

where ϵ is a quantity equal to 2 or 1 for low-pass or band-pass detectors, respectively. In these expressions W_x is the effective noise bandwidth of the signal and the two noises, ρ_1 and ρ_2 are the signal-to-noise power ratios $X(\omega)/N_1(\omega)$ and $X(\omega)/N_2(\omega)$ respectively, and F is a *spectrum form factor*,

$$F(D) = \frac{\int_0^\infty D^2(\omega)\,d\omega}{D_{\max}(\omega)\int_0^\infty D(\omega)\,d\omega} \tag{16}$$

some values of which are tabulated in Table I.[10] (D is an arbitrary power density spectrum).

Another useful situation is that in which the two noises are white with spectral densities N_{01} and N_{02} watts per radian per second; $H(\omega)$ still unity. In this case,

$$\left(\frac{S}{N}\right)_0 = \frac{W_x}{W_f}\left[\epsilon F + \frac{N_{01}}{X_{\max}} + \frac{N_{02}}{X_{\max}} + \frac{N_{01}N_{02}}{X_{\max}^2}\frac{W'}{W_x}\right]^{-1} \tag{17}$$

where $X_{\max}$ is the maximum value of the signal power density spectrum $X(\omega)$, and W' is the bandwidth of the two white (rectangular spectrum) noises, W' being assumed large enough to include all of $X(\omega)$. This expression is seen to be similar to (15) except that the input signal-to-noise ratios are redefined in terms of power *densities* where the signal power density is that at the signal spectrum peak. Also, the form factor F enters in a different fashion.

TABLE I
VARIOUS VALUES OF THE SPECTRUM FORM FACTOR F APPEARING IN THE OUTPUT-SIGNAL-TO-NOISE EXPRESSIONS IN CERTAIN SIMPLE CASES

Type of Spectrum	Density Function $D(\omega)$	Form Factor $F(D)$
Rectangular	$\begin{cases}1 \text{ for } \omega \text{ in bandwidth } \Omega \\ 0 \text{ otherwise}\end{cases}$	1
Triangular	$\begin{cases}1 - 2\lvert\omega - \omega_c\rvert/\Omega \text{ for } \lvert\omega - \omega_c\rvert < \Omega/2 \\ 0 \text{ otherwise}\end{cases}$	2/3
Gaussian	$\exp[-(\omega - \omega_c)^2]$	$1/\sqrt{2}$
Exponential	$\exp[-\lvert\omega - \omega_c\rvert]$	1/2
First-order Butterworth (single-tuned RLC circuit)	$[1 - (\omega - \omega_c)^2]^{-1}$	1/2
nth-order Butterworth	$[1 - (\omega - \omega_c)^{2n}]^{-1}$	1-1/2n

Comparison of the Two Types of Correlation Detector

A comparison of (9) and (11) shows that the main difference between the two is in the numerator and denominator $X \times Y$ terms; that is, the output signal and self-noise power, respectively. The first of these proves to be the more interesting. To bring out the difference in behavior of the two types of detector, let the "perturbing filter" be just a pure delay, τ. That is, set $H(\omega) = e^{i\omega\tau}$. Then the signal output *voltage*, the square root of the numerator of (9), is

$$\int_0^\infty X(\omega)\cos\omega\tau\,d\omega \tag{18}$$

or the correlation function, whereas the square-root of the numerator of (11) is

$$\left\{\left[\int_0^\infty X(\omega)\cos\omega\tau\,d\omega\right]^2 + \left[\int_0^\infty X(\omega)\sin\omega\tau\,d\omega\right]^2\right\}^{1/2} \tag{19}$$

or the *envelope* of the correlation function. That is, if we call the correlation function

$$\begin{aligned}\phi_x(\tau) &= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x(t)x(t+\tau)\,d\tau \\ &= \int_{-\infty}^{\infty} X(\omega)\cos\omega\tau\,d\omega = \int_0^\infty X(\omega)\cos\omega\tau\,d\omega\end{aligned} \tag{20}$$

(because of $X(\omega)$ being zero on the negative ω axis), and similarly write the Hilbert transform of the correlation function[11]

[10] The case originally treated by Fano (footnote 2), used $H(\omega) = e^{i\omega\tau}$, and employed the "single-tuned-circuit" shape of spectrum throughout. Using the appropriate value $F = \frac{1}{2}$, our (15) reduces to his corresponding results.

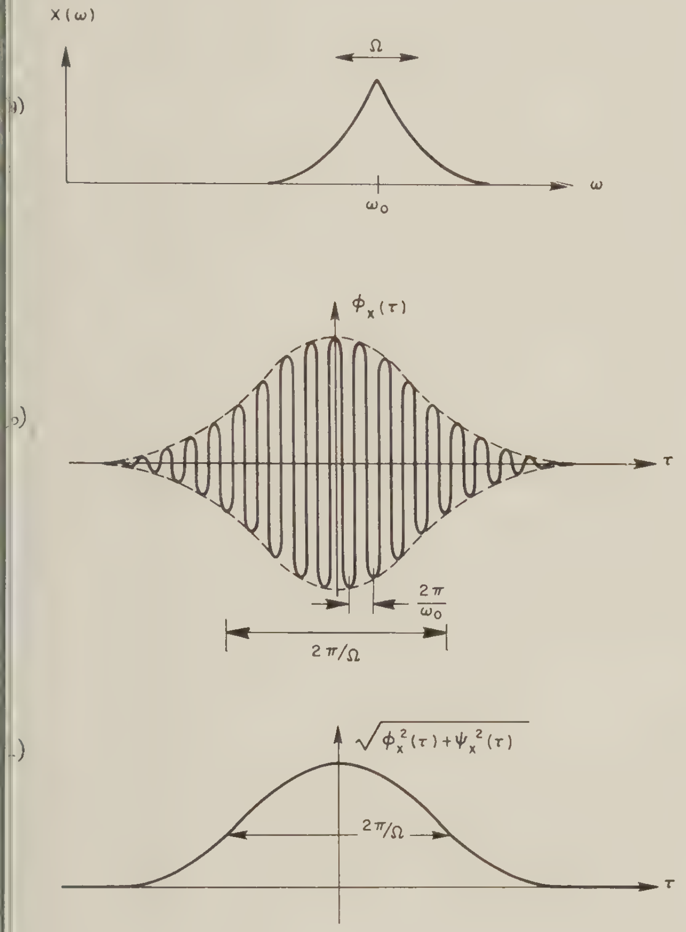
[11] One can imagine $\phi_x(\tau)$ as the projection on the real axis of a complex vector that rotates with τ at a uniform angular velocity, and $\psi_x(\tau)$ as the projection on the imaginary axis of the same vector.

$$_x(\tau) = \int_{-\infty}^{\infty} X(\omega) \sin \omega\tau \, d\omega = \int_{0}^{\infty} X(\omega) \sin \omega\tau \, d\omega, \qquad (21)$$

ıen (18) is $\phi_x(\tau)$ and (19) is

$$[\phi_x^2(\tau) + \psi_x^2(\tau)]^{1/2}. \qquad (22)$$

Fig. 4 illustrates the difference. At a is the spectral ensity of a function whose bandwidth happens to be uch smaller than its center frequency. Curve b shows ıe signal output voltage as a function of τ from a lowass correlator, and c the band-pass detector output.



ig. 4—Illustrating the action of a band-pass correlation detector in giving the envelope of the correlation function rather than the correlation function itself as with the low-pass correlation detector. (a) Typical power density spectrum, (b) output voltage of low-pass detector vs delay τ, (c) same for band-pass detector.

Conclusion

A correlation detector is defined here as a set of circuit lements that forms the integrated, or averaged, product f two time functions which differ in that one or both as noise added to it and also has been perturbed by assage through a known linear "perturbing filter" haracteristic $H(\omega)$. Expressions were derived for the utput signal-to-noise ratio of two types of correlation etector: 1) The low-pass detector, in which the integrator an arbitrary low-pass filter, (9), and 2) the band-pass orrelation detector, (11), in which one of the two input nctions was deliberately displaced in frequency by Δ, and the integrator is therefore an arbitrary band-pass filter tuned to Δ. It was assumed that signal and noises are Gaussian.[12]

The output signal-to-noise ratio is a function of the signal and noise power density spectra, and the system functions of the linear filters in the two inputs and the narrow-band integrating filter. In certain simple cases, the output signal-to-noise ratio can be expressed more easily using the effective noise bandwidth of the integrating filter and a certain spectrum form factor, [(16) and Table I]. The expressions derived for output signal-to-noise ratio extend previously available results to include arbitrary signal, noise, and integrating filter spectral shapes and the presence of the perturbing filter.

It was also shown that although the operations performed by the band-pass type of detector are considerably different from the true mathematical operation of correlation, the signal-to-noise ratio results are substantially the same, and the signal output as a function of a relative delay between input signals is the (suitably defined) envelope of the correlation function (22) rather than the correlation function itself, (18).[13]

Appendix I

Low-Pass Detector Calculations

When $u_1(t)$ and $u_2(t)$ [defined as Fourier expansions in (1) through (5)] are multiplied together (Fig. 1), there are four sets of product terms, the $X \times Y$ terms (I), the $X \times N_1$ terms (II), the $Y \times N_2$ terms (III), and the $N_1 \times N_2$ terms (IV) The multiplier output voltage at $\omega = 0$ for the first of these is

$$V_{\mathrm{I}}(0) = \frac{1}{2} \sum_{i=1}^{\theta W} \xi_i^2 h_i \cos \eta_i \qquad (23)$$

and the power is

$$P_{\mathrm{I}}(0) = V_{\mathrm{I}}^2(0) = \frac{1}{4} \sum_i \xi_i^4 h_i^2 \cos^2 \eta_i$$

$$+ \frac{1}{4} \sum_i \xi_i^2 h_i \cos \eta_i \sum_{j \neq i} \xi_j^2 h_j \cos \eta_j . \qquad (24)$$

(Summations will be assumed to run from 1 to θW.) The ensemble average power in this spectral line is

$$\overline{P_I(0)} = \frac{1}{4} \sum_i \overline{\xi_i^4} h_i^2 \cos^2 \eta_i$$

$$+ \frac{1}{4} \sum_i \overline{\xi^2} h_i \cos \eta_i \sum_{j \neq i} \overline{\xi_j^2} h_j^2 \cos \eta_j . \qquad (25)$$

The bar represents the ensemble average.

[12] Only the self-noise terms are incorrect if this assumption is violated, so long as a proportional relationship like (6) holds. (Our self-noise results depend on the statistical independence of Fourier coefficients for $i \neq j$.)

[13] The reviewer has kindly pointed out that the band-pass detector performance could be obtained with the low-pass detection scheme as follows: A second low-pass detector is constructed identical to that in Fig. 1, except that $H(\omega)$ is replaced by its complex conjugate. The output voltages of the two low-pass detectors are then combined by squaring each, adding, and then taking the square root of the sum.

For frequencies different from zero ($\omega = 2\pi n/\theta$, $n = 1, 2, \cdots \theta W$)

$$V_{\mathrm{I}}\left(\frac{2\pi n}{\theta}\right) = \frac{1}{2} \sum_i \xi_i \xi_{i+n} h_{i+n} \cos\left(\frac{2\pi n t}{\theta} + \phi_{i+n} + \eta_{i+n} - \phi_i\right) + \frac{1}{2} \sum_i \xi_{i+n} \xi_i h_i \cos\left(\frac{2\pi n t}{\theta} + \phi_{i+n} - \eta_i - \phi_i\right)$$

$$= \frac{1}{2} \cos \frac{2\pi n t}{\theta} \sum_i (\xi_i \xi_{i+n} h_{i+n} \cos A_{in} + \xi_{i+n} \xi_i h_i \cos B_{in})$$

$$- \frac{1}{2} \sin \frac{2\pi n t}{\theta} \sum_i (\xi_i \xi_{i+n} h_{i+n} \sin A_{in} + \xi_{i+n} \xi_i h_i \sin B_{in}) \quad (26)$$

using the abbreviations $A_{in} = (\phi_{i+n} + \eta_{i+n} - \phi_i)$ and $B_{in} = (\phi_{i+n} - \eta_i - \phi_i)$. The power in each of these spectral lines at nonzero frequency is one half the square of the amplitude, the latter being the sum of the squares of the two summations in (26).

$$P_{\mathrm{I}}\left(\frac{2\pi n}{\theta}\right) = \frac{1}{8} \sum [(\xi_i \xi_{i+n} h_{i+n} \cos A_{in} + \xi_{i+n} \xi_i h_i \cos B_{in})^2 + (\xi_i \xi_{i+n} h_{i+n} \sin A_{in} + \xi_{i+n} \xi_i h_i \sin B_{in})^2$$

$+$ other terms, each of which is the product of cosine or sine of A_{in} or B_{in} with cosine or sine of A_{jn} or B_{jn} with $i \neq j$. (27)

In taking the ensemble average of this expression, each of the last-named terms becomes zero. To show this, one can make a decomposition of a typical term, say $\cos A_{in}$ $\cos B_{jn}$, into a number of subterms, each of which contains $\cos \phi_{i\,\mathrm{or}\,j}$ or $\sin \phi_{i\,\mathrm{or}\,j}$ as a factor. Because the phase coefficients ϕ_i and ϕ_j ($i \neq j$) are independent, and distributed with uniform probability over the interval $(-\pi, \pi)$, these ensemble averages are zero. We thus have

$$\overline{P_{\mathrm{I}}\left(\frac{2\pi n}{\theta}\right)} = \frac{1}{8} \sum_i \overline{\xi_i^2}\, \overline{\xi_{i+n}^2} (h_i^2 + 2h_i h_{i+n} \cos (\eta_i + \eta_{i+n}) + h_{i+n}^2) \qquad n \neq 0. \quad (28)$$

Going on to the $X \times N_1$ term at zero frequency,

$$V_{\mathrm{II}}(0) = \frac{1}{2} \sum_i \nu_i \xi_i \cos (\phi_i - \gamma_i) \quad (29)$$

from which

$$\overline{P_{\mathrm{II}}(0)} = \frac{1}{4} \sum_i \overline{\nu_i^2}\, \overline{\xi_i^2}\, \overline{\cos^2 (\phi_i - \gamma_i)} + \frac{1}{4} \sum_i \overline{\nu_i}\, \overline{\xi_i}\, \overline{\cos (\phi_i - \gamma_i)} \sum_{j \neq i} \overline{\nu_j}\, \overline{\xi_j}\, \overline{\cos (\phi_j - \gamma_j)}. \quad (30)$$

But $\overline{\cos (\phi_i - \gamma_i)}$ is zero for any i, since ϕ_i and γ_i are independent ($x(t)$ and $n_1(t)$ being independent) and uniformly distributed over the interval $(-\pi, \pi)$. Therefore only the first term of (30) remains, and

$$\overline{P_{\mathrm{II}}(0)} = \frac{1}{8} \sum_i \overline{\nu_i^2}\, \overline{\xi_i^2}. \quad (31)$$

Similarly

$$\overline{P_{\mathrm{III}}(0)} = \frac{1}{8} \sum_i \overline{\mu_i^2}\, \overline{\xi_i^2}\, h_i^2, \quad (32)$$

and

$$\overline{P_{\mathrm{IV}}(0)} = \frac{1}{8} \sum \overline{\nu_i^2}\, \overline{\mu_i^2}. \quad (33)$$

As for the $X \times N_1$ term away from zero frequency, we have

$$V_{\mathrm{II}}\left(\frac{2\pi n}{\theta}\right) = \frac{1}{2} \sum_i \nu_i \xi_{i+n} \cos\left(\frac{2\pi n t}{\theta} + \phi_{i+n} - \gamma_i\right) + \nu_{i+n} \xi_i \cos\left(\frac{2\pi n t}{\theta} - \phi_i + \gamma_{i+n}\right) \quad (34)$$

from which

$$\overline{P_{\mathrm{II}}\left(\frac{2\pi n}{\theta}\right)} = \frac{1}{8} \sum_i (\overline{\nu_i^2}\, \overline{\xi_{i+n}^2} + \overline{\nu_{i+n}^2}\, \overline{\xi_i^2}) \quad (35)$$

by the same steps as those leading from (26) to (27). Similarly for the $Y \times N_2$ term

$$\overline{P_{\mathrm{III}}\left(\frac{2\pi n}{\theta}\right)} = \frac{1}{8} \sum_i (\overline{\mu_i^2}\, \overline{\xi_{i+n}^2}\, h_{i+n}^2 + \overline{\mu_{i+n}^2}\, \overline{\xi_i^2}\, h_i^2), \quad (36)$$

and the $N_1 \times N_2$ term

$$\overline{P_{\mathrm{IV}}\left(\frac{2\pi n}{\theta}\right)} = \frac{1}{8} \sum_i (\overline{\nu_i^2}\, \overline{\mu_{i+n}^2} + \overline{\nu_{i+n}^2}\, \overline{\mu_i^2}). \quad (37)$$

Now that all four components of the multiplier output spectrum have been computed, they must be operated on by the integrating filter's response $| I(\omega) |^2$. Treating the signal component [second term of (25)] in this way, and then taking the limit as $\theta \to \infty$, we have the output signal power:

$$S_0 = \lim_{\theta \to \infty} \frac{1}{4} | I(0) |^2 \sum_i \overline{\xi_i^2}\, h_i \cos \eta_i \sum_{j \neq i} \overline{\xi_j^2}\, h_j \cos \eta_j \quad (38)$$

$$= \lim_{\substack{\theta \to \infty \\ \Delta\omega = 2\pi/\theta}} | I(0) |^2 \sum_i X_i h_i \cos \eta_i \Delta\omega \sum_{j \neq i} X_j h_j \cos \eta_j \Delta\omega \quad (39)$$

from (6).

$$S_0 = | I(0) |^2 \left[\int_0^\infty X(\omega)\, | H(\omega) | \cos \eta(\omega)\, d\omega \right]^2 = | I(0) |^2 \left[\int_0^\infty X(\omega) Re(H(\omega))\, d\omega \right]^2 \quad (40)$$

where the abbreviation "Re" means "real part of."

The fluctuation power N_0 at the output can be found by multiplying $| I(2\pi n/\theta) |^2$ by all the corresponding multiplier output terms save only the second term of (25) and going to the limit as $\theta \to \infty$.

$$N_0 = \lim_{\theta\to\infty} \sum_{n=0}^{\infty} \left| I\left(\frac{2\pi n}{\theta}\right) \right|^2 \left[\overline{P_{\mathrm{I}}^{*}\left(\frac{2\pi n}{\theta}\right)} + \overline{P_{\mathrm{II}}\left(\frac{2\pi n}{\theta}\right)} + \overline{P_{\mathrm{III}}\left(\frac{2\pi n}{\theta}\right)} + \overline{P_{\mathrm{IV}}\left(\frac{2\pi n}{\theta}\right)} \right] \qquad (41)$$

where the superscript * indicates the omission of the last term when $n = 0$. A typical term in this expression, say the last one, becomes from (37),

$$\lim_{\substack{\theta\to\infty \\ \Delta\omega = 2\pi/\theta}} \frac{1}{8} \sum_{n=0}^{\infty} \left| I\left(\frac{2\pi n}{\theta}\right) \right|^2 \sum_i \left[N_1\left(\frac{2\pi i}{\theta}\right) N_2\left(\frac{2\pi(i+n)}{\theta}\right) + N_1\left(\frac{2\pi(i+n)}{\theta}\right) N_2\left(\frac{2\pi i}{\theta}\right) \right] (\Delta\omega)^2$$

$$= \frac{1}{2} \int_0^{\infty} | I(\omega) |^2 \, d\omega \int_0^{\infty} [N_1(u) N_2(u+\omega) + N_1(u+\omega) N_2(u)] \, du. \qquad (42)$$

The final signal-to-noise ratio $(S/N)_0$ is the ratio of expressions (40) and (41), namely (9).

It should be clear now how the second term in (25) was recognized as the output signal term. In using the limiting relations (6) and (7) for replacing Fourier coefficients with spectral densities, the first term of (3) becomes an integral times $\Delta\omega$ which is approaching zero. The second term, however, is an integral squared and remains as a constant spectral line at the origin during the limiting process.

Appendix II

Band-Pass Detector Calculations

The calculations for the band-pass detector are considerably simpler than for the low-pass case. When the signals $x(t)$ and $y(t)$ are multiplied, the voltage at the difference frequency Δ is

$$V_{\mathrm{I}}(\Delta) = \frac{1}{2} \sum_i \xi_i^2 h_i \cos(\Delta t + \eta_i) \qquad (43)$$

and the power is one half the squared amplitude, giving as the ensemble average,

$$\overline{P_{\mathrm{I}}(\Delta)} = \overline{\frac{1}{8}\left[\sum_i \xi_i^2 h_i \cos \eta_i\right]^2} + \overline{\frac{1}{8}\left[\sum_i \xi_i^2 h_i \sin \eta_i\right]^2}$$

$$= \frac{1}{8} \sum_i \overline{\xi_i^4} h_i^2 + \frac{1}{8} \sum_i \overline{\xi_i^2} h_i \cos \eta_i \sum_{i \neq j} \overline{\xi_j^2} h_j \cos \eta_j$$

$$+ \frac{1}{8} \sum_i \overline{\xi_i^2} h_i \sin \eta_i \sum_{i \neq j} \overline{\xi_j^2} h_j \sin \eta_j. \qquad (44)$$

For frequencies different from Δ ($\omega = \Delta + 2\pi n/\theta$ $n = \pm 1, \pm 2, \cdots \pm \theta W$)

$$V_{\mathrm{I}}\left(\Delta + \frac{2\pi n}{\theta}\right) = \frac{1}{2} \sum_i \xi_i \xi_{i+n} h_{i+n} \cos\left(\Delta t + \frac{2\pi n t}{\theta} + \phi_{i+n} - \phi_i + \eta_{i+n}\right)$$

$$= \frac{1}{2} \cos\left(\Delta t + \frac{2\pi n t}{\theta}\right) \sum_i \xi_i \xi_{i+n} h_{i+n} \cdot \cos(\phi_{i+n} - \phi_i + \eta_{i+n})$$

$$- \frac{1}{2} \sin\left(\Delta t + \frac{2\pi n t}{\theta}\right) \sum_i \xi_i \xi_{i+n} h_{i+n} \cdot \sin(\phi_{i+n} - \phi_i + \eta_{i+n}) \qquad (45)$$

from which

$$\overline{P_{\mathrm{I}}\left(\Delta + \frac{2\pi n}{\theta}\right)} = \frac{1}{8} \sum_i \overline{\xi_i^2}\, \overline{\xi_{i+n}^2} h_{i+n}^2 [\overline{\cos^2(\phi_{i+n} - \phi_i + \eta_{i+n})} + \overline{\sin^2(\phi_{i+n} - \phi_i + \eta_{i+n})}]$$

$$= \frac{1}{8} \sum_i \overline{\xi_i^2}\, \overline{\xi_{i+n}^2} h_{i+n}^2. \qquad (46)$$

In a similar fashion, we have for the $X \times N_1$ term at any frequency, including Δ,

$$V_{\mathrm{II}}\left(\Delta + \frac{2\pi n}{\theta}\right) = \frac{1}{2} \sum_i \nu_{i+n} \xi_i \cdot \cos\left(\Delta t + \frac{2\pi n t}{\theta} + \gamma_{i+n} - \phi_i\right) \qquad (47)$$

where $n = 0, \pm 1, \cdots, \pm \theta W$. From this

$$\overline{P_{\mathrm{II}}\left(\Delta + \frac{2\pi n}{\theta}\right)} = \frac{1}{8} \sum_i \overline{\nu_{i+n}^2}\, \overline{\xi_i^2}. \qquad (48)$$

The $Y \times N_2$ and $N_1 \times N_2$ terms are computed similarly:

$$\overline{P_{\mathrm{III}}\left(\Delta + \frac{2\pi n}{\theta}\right)} = \frac{1}{8} \sum_i \overline{\xi_{i+n}^2}\, \overline{\mu_i^2} h_{i+n}^2, \qquad (49)$$

$$\overline{P_{\mathrm{IV}}\left(\Delta + \frac{2\pi n}{\theta}\right)} = \frac{1}{8} \sum_i \overline{\nu_{i+n}^2}\, \overline{\mu_i^2}. \qquad (50)$$

This time, the signal power is represented by the last two of the three terms in (44). Upon multiplying this by the integrating filter response and letting $\theta \to \infty$

$$S_0 = \lim_{\substack{\theta\to\infty \\ \Delta\omega = 2\pi/\theta}} | I(\Delta) |^2 \left[\frac{1}{2} \sum_i X_i h_i \cdot \cos \eta_i \Delta\omega \sum_{j \neq i} X_j h_j \cos \eta_j \Delta\omega + \frac{1}{2} \sum_i X_i h_i \sin \eta_i \Delta\omega \sum_{j \neq i} X_j h_j \sin \eta_j \Delta\omega \right]$$

$$= \frac{1}{2} \left[\int_0^{\infty} X(\omega) Re[H(\omega)] \, d\omega \right]^2 + \frac{1}{2} \left[\int_0^{\infty} X(\omega) \, Im \, [H(\omega)] \, d\omega \right]^2. \qquad (51)$$

The output noise power is

$$N_0 = \lim_{\theta\to\infty} \sum_{n=-\infty}^{\infty} \left| I\left(\Delta + \frac{2\pi n}{\theta}\right) \right|^2 \left[\overline{P_{\mathrm{I}}^{*}\left(\Delta + \frac{2\pi n}{\theta}\right)} + \overline{P_{\mathrm{II}}\left(\Delta + \frac{2\pi n}{\theta}\right)} + \overline{P_{\mathrm{III}}\left(\Delta + \frac{2\pi n}{\theta}\right)} + \overline{P_{\mathrm{IV}}\left(\Delta + \frac{2\pi n}{\theta}\right)} \right] \quad (52)$$

where the superscript * indicates omission of the last two terms when $n = 0$.

$$N_0 = \frac{1}{2} \int_0^{\infty} \mid I(\omega) \mid \int_0^{\infty} [X(u)X(u+\omega) \mid H(u+\omega) \mid^2 + X(u)N_1(u+\omega) + X(u+\omega) \mid H(u+\omega) \mid^2 N_2(u+\Delta+\omega) + N_1(\omega)N_2(u+\Delta+\omega)]\, du\, d\omega \quad (53)$$

The ratio of (51) to (53) is the desired $(S/N)_0$, (11).

Acknowledgment

The author would like to express his gratitude to Drs. B. L. Basore, W. B. Davenport, Jr., R. M. Fano, and R. Price for much of the motivation for this study, and for many helpful suggestions.

Error Rates in Pulse Position Coding*

L. LORNE CAMPBELL†

Summary—**An expression for the error rate in a system using a binary pulse position code is derived. In the system considered, the pulses amplitude modulate a carrier and the resultant signal is contaminated by additive Gaussian noise. At the receiver the pulses are recovered by an envelope detector. If synchronization errors and post-detection filtering are neglected, it is shown that the probability of a binary error is approximated well by 1/2 exp $(-a^2/2)$, where a^2 is the peak input signal-to-noise power ratio. Finally, the error rate is derived for the case where the signal amplitude is subject to random fading. Some comparisons are made with error rates derived by Montgomery for other systems with and without carrier fading. It is found that when the signal is subject to fading the pulse position system is better than a comparable system using threshold detection.**

List of Principal Symbols

T = frame duration.

F = carrier frequency.

$w(f)$ = power spectrum of input noise.

P = pulse amplitude.

P_0 = root-mean-square pulse amplitude with Rayleigh fading.

ψ_0 = $\int_0^\infty w(f)df$ = total input noise power.

r = envelope of relative autocorrelation function of input noise for a delay $T/2$.

a^2 = $P^2/(2\psi_0)$ = input signal-to-noise power ratio at the pulse peak.

a_0^2 = $P_0^2/(2\psi_0)$ = mean signal-to-noise power ratio with Rayleigh fading.

$L_n^\alpha(x)$ = generalized Laguerre polynomial.

$I_0^{(k)}(x)$ = kth derivative of modified Bessel function of order zero.

p = probability of error in one binary digit in the pulse position system with steady signal.

p_c = probability of error in a character in a five-digit code.

p_T = probability of error in envelope detection, fixed threshold system with no fading.

p_F = probability of error in pulse position system with Rayleigh fading.

p_{FT} = probability of error in envelope detection, fixed threshold system with Rayleigh fading.

p_{FS} = probability of error in synchronous detection system with Rayleigh fading.

Introduction

BINARY pulse position coding appears to offer some advantages over a simple on-off pulse code, particularly when the signal amplitude is subject to large fluctuations. The object of this paper is to estimate the frequency of errors in a communication system using this form of coding.

In the system considered here, the time scale is divided into frames of duration T. A pulse is transmitted in either the first half or the second half of the frame. The position of the pulse in the frame then conveys one binary digit of information. In this paper, we consider a communication system in which these pulses are used to amplitude modulate a radio frequency carrier. At the receiver the pulses are regained at the output of an envelope detector. The decision as to whether the pulse is in the first or the

* Manuscript received by the PGIT, September 18, 1956.

† Radio Physics Lab., Def. Res. Board, Ottawa, Canada. Work performed under project PCC No. D48-28-35-05.

cond half of the frame is made by comparing the outputs the detector at the times $T/4$ and $3T/4$, measured om the beginning of the frame. The pulse is then said be in the first or second position according as the output the time $T/4$ or $3T/4$ is greater.

Errors may occur if the transmission link is noisy. The fect of additive white Gaussian noise will be considered ere. The spectrum of the noise at the input to the detector then determined by the predetection filters in the ceiver. It will also be assumed here that the receiver nows when each frame begins. This information may be rovided on a parallel synchronization channel or by ther means.

Derivation of Error Rate—Uncorrelated Noise

Let the input to the detector be

$$V(t) = \begin{cases} P\cos 2\pi Ft + V_n(t) & \text{for} \quad 0 < t < T/2, \\ V_n(t) & \text{for} \quad T/2 < t < T. \end{cases} \tag{1}$$

n (1), $V_n(t)$ is a narrow-band Gaussian noise voltage ith power spectrum $w(f)$, and $P\cos 2\pi Ft$, for $< t < T/2$, is a square pulse of amplitude P. This nput is what we would have if the pulse is transmitted in he first half of the frame $0 < t < T$. If the pulse were in the econd half the analysis would be exactly the same. The nalysis would not be affected if the pulse is shaped, rovided that the pulse amplitude at $t = T/4$ is P and t $t = 3T/4$ is zero.

Let the output of the envelope detector be $R(t)$. Then $R(T/4) > R(3T/4)$ we say that the pulse is in the rst half-frame and conversely. Thus, for the input (1), n error is made if $R(T/4) < R(3T/4)$.

If the noise voltage, $V_n(t)$, is confined to a spectral and whose width is small compared with F we can write

$$V(t) = \begin{cases} [A(t) + P]\cos 2\pi Ft \\ \qquad + B(t)\sin 2\pi Ft & (0 < t < T/2), \\ A(t)\cos 2\pi Ft \\ \qquad + B(t)\sin 2\pi Ft & (T/2 < t < T), \end{cases} \tag{2}$$

here $A(t)$ and $B(t)$ are normally distributed and are lowly varying compared with $\cos 2\pi Ft$. The output of he detector is then

$$R(t) = \begin{cases} [(A+P)^2 + B^2]^{1/2} & (0 < t < T/2), \\ [A^2 + B^2]^{1/2} & (T/2 < t < T). \end{cases} \tag{3}$$

We shall first derive an expression for the error rate on he assumption that the input noise at $t = T/4$ is effecively uncorrelated with the input noise at $t = 3T/4$. Let

$$R_1 = R(T/4), \qquad R_2 = R(3T/4). \tag{4}$$

According to the well-known result on the output of nvelope detectors,[1] the probability density of R_1 is

$$\frac{R_1}{\psi_0}\exp\left[-\frac{(R_1^2+P^2)}{2\psi_0}\right]I_0\left(\frac{R_1P}{\psi_0}\right)$$

and the probability density of R_2 is

$$\frac{R_2}{\psi_0}\exp\left(\frac{-R_2^2}{2\psi_0}\right),$$

where ψ_0, the total input noise power, is given by

$$\psi_0 = \int_0^\infty w(f)\,df. \tag{5}$$

Since, by hypothesis, R_1 and R_2 are uncorrelated, the joint probability density of R_1 and R_2, $F(R_1, R_2)$, is given by

$$F(R_1, R_2) = \frac{R_1R_2}{\psi_0^2}\exp\left[-\frac{(R_1^2+R_2^2+P^2)}{2\psi_0}\right]I_0\left(\frac{R_1P}{\psi_0}\right). \tag{6}$$

Since an error occurs if $R_2 > R_1$, the probability of an error in a frame, p, is given by

$$p = \int_0^\infty dR_1 \int_{R_1}^\infty dR_2\, F(R_1, R_2). \tag{7}$$

This integral is easily evaluated [see (51) in the Appendix] to give

$$p = \frac{1}{2}\exp\left(\frac{-a^2}{2}\right), \tag{8}$$

where a^2 is the signal-to-noise power ratio in the input at $t = T/4$. That is,

$$a^2 = P^2/2\psi_0. \tag{9}$$

Eq. (8) gives a simple approximation to the probability of error in a frame on the assumption that the noise at the two sampling positions is uncorrelated. It will be shown later that (8) remains a good approximation in many cases in which correlation might be expected to play a part.

Derivation of Error Rate—Correlated Noise

A more accurate expression for the error probability is obtained if we take account of the correlation between R_1 and R_2. To this end we let

$$x_1 = P + A(T/4), \qquad y_1 = B(T/4)$$
$$x_2 = A(3T/4), \qquad y_2 = B(3T/4).$$

Then, by a simple modification of the result given by Rice[2] for the case of no signal, the joint probability density function of x_1, x_2, y_1, y_2 is given by

$$\begin{aligned} p_4(x_1, x_2, y_1, y_2) &= [4\pi^2(\psi_0^2 - \mu_{13}^2 - \mu_{14}^2)]^{-1} \\ &\cdot \exp - \{\psi_0[(x_1-P)^2 + x_2^2 + y_1^2 + y_2^2] \\ &- 2\mu_{13}[(x_1-P)x_2 + y_1y_2] \\ &- 2\mu_{14}[(x_1-P)y_2 - y_1x_2]\}/2(\psi_0^2 - \mu_{13}^2 - \mu_{14}^2), \end{aligned} \tag{10}$$

[1] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 24, pp. 46–156; January, 1945. See section 3.10.

[2] *Ibid.*, section 3.7.

where

$$\mu_{13} = \int_0^\infty w(f) \cos \pi(f - F)T \, df, \tag{11}$$

and

$$\mu_{14} = \int_0^\infty w(f) \sin \pi(f - F)T \, df. \tag{12}$$

We now make the substitutions

$$x_i = R_i \cos \theta_i, \qquad y_i = R_i \sin \theta_i \qquad (i = 1, 2).$$

Then R_1 and R_2 represent the detector outputs at the times $T/4$ and $3T/4$ respectively. R_1 and R_2 have the joint probability density function $F(R_1, R_2)$, where

$$F(R_1, R_2) = R_1 R_2 \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2$$

$$\cdot p_4(R_1 \cos \theta_1, R_2 \cos \theta_2, R_1 \sin \theta_1, R_2 \sin \theta_2). \tag{13}$$

Since an error occurs if $R_2 > R_1$, the probability of error, p, is given, as before, by

$$p = \int_0^\infty dR_1 \int_{R_1}^\infty dR_2 \, F(R_1, R_2). \tag{7}$$

The details of the calculation of p are left to the Appendix. The result is that

$$p = \sum_{n=0}^\infty f_n r^{2n}, \tag{14}$$

where

$$r = \frac{(\mu_{13}^2 + \mu_{14}^2)^{1/2}}{\psi_0}, \tag{15}$$

and

$$f_n = 2e^{-a^2} \int_0^\infty \rho e^{-2\rho^2} [L_n(\rho^2) - L_{n-1}(\rho^2)]$$

$$\cdot \sum_{k\;0}^{n} L_{n-k}^k (a^2 + \rho^2) \frac{(2a\rho)^k}{k!} I_0^{(k)} (2a\rho) \, d\rho. \tag{16}$$

The functions $L_n^\alpha(x)$ are the generalized Laguerre polynomials and are defined by

$$L_n^\alpha(x) = \frac{e^x}{n!} x^{-\alpha} \frac{d^n}{dx^n} (e^{-x} x^{n+\alpha}). \tag{17}$$

$L_n^0(x)$ will often be written $L_n(x)$. Also, in (16),

$$I_0^{(k)}(x) = \frac{d^k}{dx^k} I_0(x), \tag{18}$$

where $I_0(x)$ is the modified Bessel Function of order zero.

The expansion (14) for p holds for $|r| < 1$. Ordinarily, r will be quite small compared with unity and a small number of terms of the series gives a satisfactory approximation. The quantity r is related closely to the autocorrelation function of the noise and, in general, decreases with increasing T.

The first two coefficients of the series (14), f_0 and f_1, are given by

$$f_0 = 1/2 \exp\left(\frac{-a^2}{2}\right) \tag{19}$$

and

$$f_1 = \frac{(a^4 - 4a^2)}{32} \exp\left(\frac{-a^2}{2}\right). \tag{20}$$

As shown in the Appendix, an upper bound to the absolute values of f_n can be found. This permits us to estimate the accuracy of the value of p which is obtained from the first few terms of the expansion (14). The bound is given by inequality

$$|f_n| \leq 2 \exp\left(\frac{-a^2}{4}\right). \tag{21}$$

From (14), (19), (20), and (21), we have

$$p = 1/2 \, [1 + \frac{r^2 a^2}{16} (a^2 - 4)] \exp\left(\frac{-a^2}{2}\right) + O(r^4), \tag{22}$$

where

$$|O(r^4)| \leq \sum_{n=2}^\infty 2r^{2n} \exp\left(\frac{-a^2}{4}\right) = \frac{2r^4}{1 - r^2} \exp\left(\frac{-a^2}{4}\right). \tag{23}$$

When $r = 0$, (22) reduces to the simpler (8).

Discussion of Results

In order to appreciate these results it is necessary to have some idea of the magnitude of the parameter r. We note first that r is the envelope of the relative autocorrelation function of the input noise evaluated for a delay of $T/2$. That is,

$$\frac{\psi\left(\frac{T}{2}\right)}{\psi_0} = r \cos(\pi F T + \varphi), \tag{24}$$

where $\psi(\tau)$ is the autocorrelation function of the input noise. This autocorrelation function is given by

$$\psi(\tau) = \int_0^\infty w(f) \cos 2\pi f \tau \, df. \tag{25}$$

For example, let us assume that the input noise has a rectangular spectrum centered on the frequency F. That is,

$$w(f) = \begin{cases} 0 & \text{for} \quad 0 \leq f < F - \frac{W}{2}, \\ N_0 & \text{for} \quad F - \frac{W}{2} \leq f \leq F + \frac{W}{2}, \\ 0 & \text{for} \quad f > F + \frac{W}{2}. \end{cases} \tag{26}$$

If we substitute these values in (5), (11), and (12) we have

$$\psi_0 = N_0 W, \qquad \mu_{13} = \frac{2N_0}{\pi T} \sin \frac{\pi T W}{2}, \qquad \mu_{14} = 0. \tag{27}$$

us

$$a^2 = \frac{P^2}{2N_0W} \tag{28}$$

l

$$r = \frac{2}{\pi TW} \sin \frac{\pi TW}{2}. \tag{29}$$

Now the frame frequency is T^{-1}. Since the pulses are nfined to a duration which is closer to $T/2$, the signal nction will have a strong harmonic component of quency $2T^{-1}$. Since this signal amplitude modulates a rrier, the input bandwidth must be at least $4T^{-1}$ if important sidebands are to be included. If the pulses not well shaped a larger bandwidth might be necessary. us the magnitude of r will usually be $(2\pi)^{-1}$ ($\doteqdot 0.159$) less. Calculations with other shapes of input spectra licate that r will usually be much less than this.

When $r = 0.16$, the simple expression

$$p = 1/2 \exp\left(\frac{-a^2}{2}\right) \tag{8}$$

an excellent approximation to the probability of error r $a^2 < 10$. The approximation remains quite good even r larger values of a.

The probability, p, of error in a frame is plotted against e peak signal-to-noise ratio, a^2, for $r = 0$ in Fig. 1. For e range of values of a in this figure the curve for $r = 0.16$ practically indistinguishable from the curve shown. he curve shows that a signal-to-noise ratio of a little ore than 10 db will give a satisfactory error rate.

It seems likely that if some form of post-detection tering or integration is introduced before the decision to the position of the pulse is made, the probability of ror should be decreased. In principle, the error probbility with arbitrary post-detection filtering could be btained from the general results of Meyer and Middlen.[3] However, the problem of solving the associated tegral equations seems to be difficult except in the mple case which is treated more directly in the present aper.

Finally, it should be pointed out that p is the probbility of error in a single frame. In a five-digit code, e probability of error in a character, p_c, is given by

$$p_c = 1 - (1 - p)^5. \tag{30}$$

hen p is small this becomes

$$p_c \doteqdot 5p. \tag{31}$$

Error Rate with a Fading Signal

In this section we present a brief discussion of the rror rate of a system using pulse position coding when the signal amplitude is subject to fluctuations. It seems clear that a detection system of this type should be better than a system which uses threshold detection of single pulses when the signal is fading.

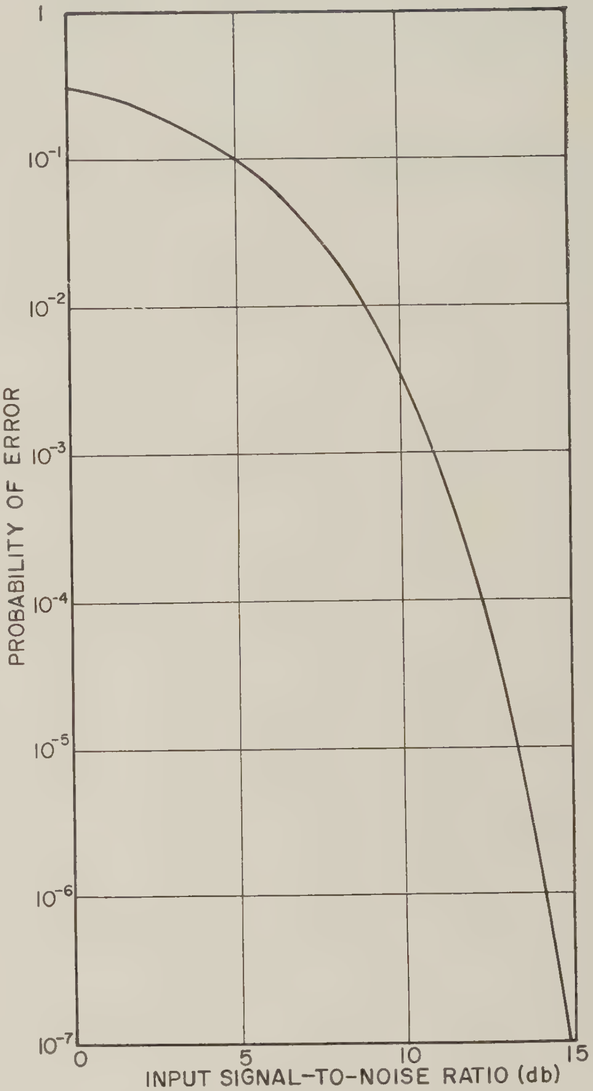


Fig. 1—The probability of error as a function of input signal-to-noise ratio in a pulse position system. No carrier fading.

Montgomery[4] has calculated the error rates for several other methods of modulation and binary coding when the signal is fading. Like Montgomery, we assume that the values of signal amplitude, P, follow a Rayleigh distribution. Thus, we suppose that the probability density function of the signal amplitude, P, is

$$\frac{2P}{P_0^2} \exp\left(\frac{-P^2}{P_0^2}\right),$$

where P_0 is the long term root-mean-square value of the signal amplitude. We shall also assume that, for given P,

[3] M. A. Meyer and D. Middleton, "On the distribution of signals nd noise after rectification and filtering," *J. Appl. Phys.*, vol. 25, p. 1037–1052; August, 1954.

[4] G. F. Montgomery, "A comparison of amplitude and angle modulation for narrow-band communication of binary-coded messages in fluctuation noise," Proc. IRE., vol. 42, pp. 447–454; February, 1954.

the error probability with pulse position coding is

$$1/2 \exp\left(\frac{-P^2}{4\psi_0}\right).$$

Then the probability of error with a fading signal, p_F, is given by

$$p_F = \int_0^\infty 1/2 \exp\left(\frac{-P^2}{4\psi_0}\right)\frac{2P}{P_0^2}\exp\left(\frac{-P^2}{P_0^2}\right)dP$$

$$= \frac{1}{2 + P_0^2/2\psi_0}. \qquad (32)$$

The error in p_F due to neglecting the remaining terms in the series (14) is easily shown to be less than

$$\frac{8r^2}{(1 - r^2)(4 + P_0^2/2\psi_0)}.$$

For $r = 0.16$, which is about the maximum possible value of r, this amounts to about 20 per cent of p_F.

Comparison with Other Systems

It is interesting to compare the results for pulse position coding, with and without a fading carrier, with the results obtained by Montgomery[4] for threshold detection of a binary coded message. In Montgomery's calculation one binary digit is transmitted by transmitting or not transmitting a pulse. The pulse is used to amplitude modulate a carrier and is recovered in the receiver at the output of an envelope detector. Such a system requires only half the bandwidth of the pulse position code and a comparison of the systems must take account of this fact.

First, we consider the case of no carrier fading. The error rate in the pulse position system is, approximately,

$$p = 1/2 \exp(-a^2/2), \qquad (8)$$

where a^2 is the signal-to-noise power ratio at the peak of the pulse. The error rate for threshold detection as given by Montgomery is, approximately,

$$p_T = 1/2 \exp\left(\frac{-a^2}{2}\right)$$

$$+ 1/4 \operatorname{erf}\frac{a}{\sqrt{2}}\left[\operatorname{erf}\frac{3a}{\sqrt{2}} - \operatorname{erf}\frac{a}{\sqrt{2}}\right], \qquad (33)$$

where erf x is the error function, defined by

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}\,du.$$

In (33) the parameter a^2 is one half the signal-to-noise ratio at the peak of a pulse. The factor one half is inserted because a transmitter power which gives a signal-to-noise ratio of a^2 in the pulse position system will give a signal-to-noise ratio of $2a^2$ in the other system because of the different bandwidths required. Thus, assuming that each system uses the least possible bandwidth, equal values of a in (8) and (33) correspond to equal values of transmitter power. It will be seen from (8) and (33) that the pulse position system gives a slightly lower error rate. The difference is significant only at low signal-to-noise ratios. Moreover, as Montgomery points out, the assumptions used in deriving (33) lead to a value of p_T which is too high at low signal-to-noise ratios. Hence there does not appear to be any practical difference between the error rates for the two systems.

We now consider the case of carrier fading. The corresponding error probability for the pulse position system is

$$p_F = \frac{1}{2 + a_0^2}, \qquad (32)$$

where a_0^2 is the mean signal-to-noise power ratio at the pulse peak, $P_0^2/(2\psi_0)$. For the threshold detection system with envelope detection, the error probability, p_{FT}, is

$$p_{FT} = 1/2\left[1 - \frac{2a_0^2}{(1 + 2a_0^2)^{(1+1/2a_0^2)}}\right]. \qquad (34)$$

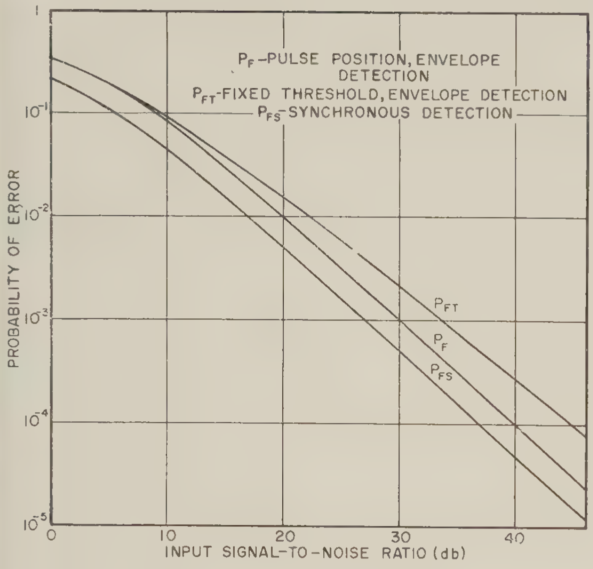
It is interesting to compare these probabilities with the error probability given by Montgomery[4] for a synchronous detector and a fading carrier. In this case, the system uses a synchronous detector with automatic gain control to maintain the threshold-to-signal ratio at its optimum value. The error probability for this system, p_{FS}, is given by

$$p_{FS} = 1/2\left[1 - \sqrt{\frac{a_0^2}{a_0^2 + 2}}\right]. \qquad (35)$$

The parameter a_0 has again been adjusted so that (32), (34), and (35) may be compared directly.

The error probabilities p_F, p_{FT}, and p_{FS} are plotted in Fig. 2, on the next page. It will be seen that p_F falls between p_{FT} and p_{FS} for the whole useful range. For mean signal-to-noise ratios of more than 30 db (binary error probability less than 10^{-3}) the error rate with the pulse position system is less than half that with the fixed-threshold, envelope-detection system, and is approximately twice the error rate of the synchronous system.

A similar comparison can be performed for the case where synchronous detection is used in place of envelope detection in the pulse position system. In this case, the inputs at carrier peaks are compared at the two pulse positions. As with the envelope detection system, the pulse is said to be in the position in which the input is greater. Without the operation of envelope detection the mathematics is much simpler. The calculations will not be reproduced here because they are very nearly the same as those performed by Montgomery for the synchronous detection of an amplitude modulated carrier. The result is that the two systems perform equally well, both with and without a fading carrier. That is, for the same error rate, an on-off amplitude modulation system, with synchronous detection and the optimum threshold setting, requires twice the signal-to-noise ratio and one half the



ig. 2—The probability of error in three systems as a function of average input signal-to-noise ratio in the pulse position system. Carrier fading.

andwidth of the pulse position system with synchronous etection. These statements are exactly true only when the oise at the two pulse positions in a frame is uncorrelated. Iowever, the correlation is usually small enough so that he statements are very nearly accurate. It should also e mentioned that the pulse position system does not equire good automatic gain control such as that required n the threshold detection system.

Figs. 1 and 2 can be used in conjunction with Montomery's curves to obtain a comparison of the pulse osition system described here with the other modulation nethods (frequency and phase modulation) examined y Montgomery. It should be remembered, however, hat these modulation methods could also be used with ulse position coding in place of the amplitude modulation onsidered here.

Appendix

In this appendix we shall outline the main steps in erforming the integrations indicated in (7) and (13). Before performing the integration in (13) it is convenient o introduce the new symbols

$$\rho_i = \frac{R_i}{\sqrt{2\psi_0}} \qquad (i = 1, 2), \tag{36}$$

$$a = \frac{P}{\sqrt{2\psi_0}}, \tag{9}$$

$$r = \frac{(\mu_{13}^2 + \mu_{14}^2)^{1/2}}{\psi_0}, \tag{15}$$

$$F_1(\rho_1, \rho_2) = 2\psi_0 F(\sqrt{2\psi_0}\,\rho_1, \sqrt{2\psi_0}\,\rho_2). \tag{37}$$

Clearly $R_1 > R_2$ if $\rho_1 > \rho_2$, and hence

$$p = \int_0^\infty d\rho_1 \int_{\rho_1}^\infty d\rho_2\, F_1(\rho_1, \rho_2). \tag{38}$$

From (10), (13), and (37)

$$F_1(\rho_1, \rho_2) = \int_0^{2\pi} \int_0^{2\pi} \frac{\rho_1\rho_2}{\pi^2(1 - r^2)} \exp\left\{\frac{-1}{1 - r^2} \cdot \left[\rho_1^2 + \rho_2^2 + a^2 - 2a\rho_1 \cos\theta_1 - \frac{2\mu_{13}}{\psi_0}\Big(\rho_1\rho_2 \cos(\theta_2 - \theta_1) - a\rho_2 \cos\theta_2\Big) - \frac{2\mu_{14}}{\psi_0}\Big(\rho_1\rho_2 \sin(\theta_2 - \theta_1) - a\rho_2 \sin\theta_2\Big)\right]\right\} d\theta_2\, d\theta_1. \tag{39}$$

The integration with respect to θ_2 is performed with the aid of the integral representation

$$I_0(x) = \frac{1}{2\pi}\int_\beta^{\beta+2\pi} e^{\pm x\cos\varphi}\, d\varphi \tag{40}$$

where β has any value whatsoever and $I_0(x)$ is the modified Bessel function of the first kind and order zero. Then

$$F_1(\rho_1, \rho_2) = \int_0^{2\pi} \frac{2\rho_1\rho_2}{\pi(1 - r^2)} \exp\left[\frac{-1}{1 - r^2}(\rho_1^2 + \rho_2^2 + a^2 - 2a\rho_1\cos\theta_1)\right] I_0(q)\, d\theta_1, \tag{41}$$

where

$$q = \frac{2\rho_2 r}{1 - r^2}(\rho_1^2 + a^2 - 2a\rho_1\cos\theta_1)^{1/2}. \tag{42}$$

The integrand in (41) is now expanded in a series of ascending powers of r. The expansion is given by the following bilinear generating function[5] for the Laguerre polynomials:

$$(1 - z)^{-1} \exp\left[-z\frac{(x + y)}{1 - z}\right](xyz)^{-\alpha/2} I_\alpha\left[\frac{2(xyz)^{1/2}}{1 - z}\right] = \sum_{n=0}^{\infty} \frac{n!}{\Gamma(n + \alpha + 1)} L_n^\alpha(x) L_n^\alpha(y) z^n \ (|z| < 1). \tag{43}$$

The Laguerre polynomials, $L_n^\alpha(x)$, are defined by (17). If we write

$$p = \sum_{n=0}^{\infty} f_n r^{2n} \tag{14}$$

and use (38), (41), and (43), we have

$$f_n = \int_0^\infty d\rho_1 \int_{\rho_1}^\infty d\rho_2 \int_0^{2\pi} d\theta_1 \frac{2\rho_1\rho_2}{\pi} \cdot \exp\left[-(\rho_1^2 + \rho_2^2 + a^2 - 2a\rho_1\cos\theta_1)\right] \cdot L_n(\rho_2^2) L_n(\rho_1^2 + a^2 - 2a\rho_1\cos\theta_1). \tag{44}$$

[5] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, "Higher Transcendental Functions," McGraw Hill Book Co., Inc., New York, N.Y., vol. 2; 1953. See eq. 10.12 (20).

Now the integration with respect to ρ_2 can be performed directly. We obtain[6]

$$2\int_{\rho_1}^{\infty} \rho_2 L_n(\rho_2^2)e^{-\rho_2{}^2}\, d\rho_2 = e^{-\rho_1{}^2}[L_n(\rho_1^2) - L_{n-1}(\rho_1^2)]. \tag{45}$$

The other Laguerre function is expanded as a power series in $\cos\theta_1$, giving

$$L_n(\rho_1^2 + a^2 - 2a\rho_1 \cos\theta_1) = \sum_{k=0}^{n} L_{n\ k}^{k}(a^2 + \rho_1^2)\frac{(2a\rho_1 \cos\theta_1)^k}{k!}. \tag{46}$$

Also, if (40) is differentiated k times, we have

$$I_0^{(k)}(x) = \frac{d^k I_0}{dx^k} = \frac{1}{2\pi}\int_0^{2\pi} e^{x\cos\varphi}\cos^k\varphi\, d\varphi. \tag{47}$$

Thus, from (46) and (47),

$$\frac{1}{2\pi}\int_0^{2\pi} e^{2a\rho_1\cos\theta_1} L_n(\rho_1^2 + a^2 - 2a\rho_1\cos\theta_1)\, d\theta_1 = \sum_{k=0}^{n} L_{n-k}^{k}(a^2 + \rho_1^2)\frac{(2a\rho_1)^k}{k!} I_0^{(k)}(2a\rho_1). \tag{48}$$

The combination of (44), (45), and (48) gives the expression (16) for f_n.

We can also obtain a bound for f_n from (44). An upper bound for the Laguerre polynomials is given[7] by

$$|\,L_n(x)\,| \leq e^{x/2} \qquad (x \geq 0). \tag{49}$$

When this is substituted in (44), the result is

$$|\,f_n\,| \leq 4e^{-a^2/2}\int_0^{\infty} \rho_1 e^{-\rho_1{}^2} I_0(a\rho_1)\, d\rho_1 = 2e^{-a^2/4}. \tag{50}$$

The last integral was obtained from Weber's first exponential integral.[8]

Finally, the coefficients f_n can always be evaluated with the aid of Hankel's formula:[9]

$$\int_0^{\infty} I_\nu\,(at)\exp(-p^2t^2)t^{\mu-1}\,dt = \frac{\Gamma(1/2[\nu+\mu])(a/2p)^\nu}{2p^\mu\Gamma(\nu+1)} \cdot {}_1F_1(1/2[\nu+\mu];\nu+1;a^2/4p^2), \tag{51}$$

where ${}_1F_1$ is a confluent hypergeometric function and a and p are arbitrary constants.

[6] *Ibid.*, eq. 10.12 (29).

[7] *Ibid.*, inequality 10.18 (3).

[8] G. N. Watson, "Theory of Bessel Functions," Cambridge University Press, Cambridge, England, p. 393; 1948.

[9] *Ibid.*

The Part of Statistical Considerations in the Separation of a Signal Masked by a Noise*

JEAN A. VILLE†

Summary—**The object of this paper is to demonstrate that the stochastic considerations presently involved in signal detections are purely descriptive and are not sufficiently developed to reach the proposed aim.**

Introduction

THIS PAPER intends to show which processes are at presently available to attempt to separate a signal masked by a noise. These processes have to be defined theoretically in order to be realized. According to the literature, these processes appear to be very numerous, but in fact, there is only a small number of them.

The optimum filter for given structures of the noise and signal is satisfactorily defined, whether the adopted solution appeals to the theories of Wiener (reduction of the standard deviation), to Fisher (estimation of the maximum likelihood), to Neymann (adjustment of the probabilities of errors of first and second type), or to Bayes (determination of the probabilities *a posteriori*). All these theories may be checked practically by considering an appropriate risk-function.

For the realization, it is necessary to know: 1) The statistical definition of the noise and signal; 2) the analytical determination of the optimum-filtered signal; and 3) the physical realization of this analytical determination.

The most important work made up to this day concerns the second class of researches.

We have attempted to show that it is the first and the third classes which actually determine the second one.

The Fundamental Operations

To emphasize the philosophy of separation of a signal masked by noise, we have to deal with two sets of signals:

* Manuscript received by the PGIT, June 1, 1956.

† Societe Alsacienne de Const. Mecaniques, Paris 13, France.

A set of pure signals $s_1, s_2, \cdots, s_n$; and 2) a set of rupted signals $S_1, S_2, \cdots, S_m$.

The signals to be separated are the signals s_i; the signals our disposal are the signals S_j; the operation of separa- n between signal and noise consists in a transformation sociating to each signal S_j one, and only one, signal s_i. is operation will be called *filtering*.

Filtering is a transformation with unique output; er reception of a corrupted signal, a regenerative cuit can only give one signal, which is, rightly or ongly, treated as being the best pure signal. If the tering is done in several steps, or if the regenerated nal has to be retransmitted, it may not be necessary useful to ask for one signal; it may be asked that a nal S_j initiates several prefiltered intermediate signals $_1, S'_{j2}, \cdots$, a remixing of which can be performed in e final filtering, or which are the inputs of several insmission lines, the outputs of which are mixed to ve a final signal. We limit our considerations to a final tering, the output of which is an unique signal.

The performance of a filter is bounded to a comparison tween the filtered signal

$$FS. \qquad (1)$$

ad the pure signal which originated S. The change of s riginal signal) into S can be symbolized by an operator (transmission) which is generally a many-to-many lation

$$S = Ts. \qquad (2)$$

The filtering problem is then, T being known, to find so that FT, product of the two operators, shall be operator very like the unity (or neutral) operator:

$$FT \approxeq 1. \qquad (3)$$

The object space of operator T is the space of pure gnals; its image space is the space of corrupted signals. he object space of F is the same as the image space of T. he image space of F would ideally be the space of pure gnals. Physically, the image space of F can only be the ace of possible outputs of the filtering device used to ealize the operator F. This space is *a priori* unknown, F eing the object of the research. Were it possible to btain *exactly* the equality (3), it would be an absolute mposed condition that the image space of F should be he same as the object space of T, *i.e.*, the space of pure ignals.

The equality (3) being only an approximate one, it s not necessary to postulate the identity of the output f F with the pure signals.

The Risk-Function

If the regeneration is only approximate, we have to ace three categories of signals: 1) The pure signals s_i;) the corrupted signals S_j; and 3) the regenerated signals $_h$, and two transformations: The transformation T of ransmission; and the transformation F of regeneration, nd a "norm," *i.e.*, an evaluation of the damage consequent to a false interpretation of S_j. This norm can be a functional of two signals s_i and σ_h:

$$R_{ih} = R(s_i, \sigma_h).$$

In all these definitions, nothing is postulated about linearity of T, F, and about the stochastic characteristics of the correspondence T between pure and corrupted signals.

The correspondence T may become a one-to-one correspondence; the regeneration can be impossible because this one-to-one correspondence is impossible to reverse physically. This is the case when T is a transmission on a delay line without distortion. The best we can do is to take

$$\sigma(t) = s(t - \theta)$$

θ being the delay time of the line. In general, it is supposed that a pure delay causes no damage at all, so that

$$R[s(t), s(t - \theta)] = 0.$$

When T is a many-to-many correspondence, the theoretical limit to an ideal regeneration is not the fact that one s_i generates several S_j, but the fact that one S_j is generated by several s_i. Concerning the σ_h, we may suppose that F is a many-to-one correspondence. This is not in fact the case, any regenerative device having its own noise, so that to one input signal are associated several output signals; but we can look at this local *filtering noise* as a secondary phenomenon.

The correspondence between the S_j and the σ_h being a many-to-one correspondence, the choice of F, if the space of the σ_h is given, is only a matter of dividing the space of S_j in disjoint subspaces, and assigning to every one of these subspaces a certain σ_h.

Why Stochastic Considerations Enter

Being given the s_i, the S_j, the σ_h, T, and the matrix $[R_{ih}]$, if we are absolutely free in the choice of F, we have to minimize a certain quantity, depending on F. This quantity concerns the complete set of s_i and S_j; for a precise definition of it, we need further conventions. There, for the first time, stochastic considerations appear. If we consider *all* the binary relations included in the correspondence T, we find in general that, for every choice of F, there exists for a given s_i a possible σ_h for which the damage R_{ih} is very severe. For continuous processes it may appear, more drastically, that for any s any σ is possible; in this situation, any regeneration is impossible. If we look at the situation from a probabilistic point of view, we have at our disposal: 1) A weighting of the s_i; and 2) a weighting of the binary relations between the s_i and the S_j, which, considered as a whole, frame the transformation T.

With these weightings, it is possible, for every many-to-one correspondence F, to define an average value of R_{ih}. If we call p_i and p_{ij} the weightings 1) and 2), and if $h(j)$ is the rank of the σ_h associated to S_j, the average value of the R_{ij} is:

$$\bar{R}_{ij} = \sum_{ij} p_i \, p_{ij} R[i, h(j)]$$

and the "best" F is the F or one of the several F's which minimize $\bar{R}$. The transformations F being defined when the function $h(j)$ is itself defined, the minimization of R is a pure algebraic problem. The difficulties are: 1) The choice of the p_i; 2) the choice of the p_{ij}; 3) the choice of the σ_h; 4) the choice of the "function of risk" R_{ih}; and 5) the realization of the correspondence $h(j)$.

Physical Limits for F

Of the difficulties 1) to 5), the main one is the last. Not all correspondences between S_j and σ_h are physically realizable. There appears the fundamental intervention of physical time. If the pure signals s_i are emitted in the interval 0, T_1, if the corrupted signals S_j are received in the interval (T_2, T_3), and if the regenerated signals σ_h are emitted in the interval (T_4, T_5), the filtering device, in the general case, can work only if $T_4 \geq T_3$; *i.e.*, the device has to store the whole signal S_j before being able to emit the signal σ_h.

If the method of separation has to be extended to indefinite signals, the storage of the whole signal S_j before emitting σ_h is unthinkable; we must restrict the field of realizable F's to the transformations which, to emit the signal $\sigma_h(\theta)$ (considered as a function of time), store only the values of $S_j(t)$ for $t < \theta$. If the signal $\sigma(\theta)$ for $t < \theta$ is compared with the signal $s(t)$ for $t < \theta - \tau$, τ is the time allowed for transmission and for the working of the filtering device. It appears that, the longer τ is, the less is the restriction imposed on F.

If the signals are pulsed signals, the physical time restriction is not drastic, theoretically, so long as a delay in the filtering is not damaging. But very strong restrictions are imposed by the technical impossibility of including a long storage in a physical filter. The only storage which is reasonably possible is a storage of energy, or any other characteristic available by integration.

In some cases, a conventional filter can provide the best solution. For instance, if $s(t)$ is a filtered Gaussian white noise, and if $S(t)$ is a sum:

$$S(t) = s(t) + b(t)$$

when $b(t)$ is another filtered Gaussian white noise, it is possible to show that a linear filtering can minimize R, if we measure $\bar{R}$ as a quadratic error. The results agree with the classical notions of signal/noise ratio. The linearity of the filtering results from the fact that the search for a minimum concerns quadratic forms, leading to linear equations; the presence of the quadratic forms is implied by the Gaussian expression of the probability laws. On Fig. 1, we see that the filter F operates on a mixture of outputs of two linear filters; it is itself a linear filter. The minimization of the conventional output of R is more and more perfect when the delay of the line between the outputs of F_s (pure signal) and F (regenerated signal) becomes longer.

When the signal, or the noise, are not Gaussian, so that F_1 or F_2 *are not* linear filters in our conventional diagram, F is itself no more linear. Suppose for instance that b being Gaussian, $s(t)$ is a filtered series of pulses emitted with a time-rule following the Poisson's Law. To extract these pulses from white noise, we have to set some thresholds, and to introduce some limiting devices. F_s being no more a linear filter, we expect to find in F some nonlinear characteristics. And this is in fact the case.

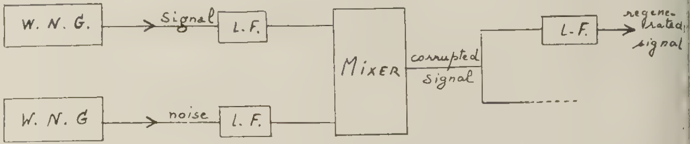


Fig. 1—The outputs of two white noise generators are filtered (by linear filters), mixed, and separated by linear filters. W. N. G. = white noise generators; L. F. = linear filter.

Stochastic Descriptions

Let us return to the different choices 1) to 4). The first two choices are a theoretical matter, interesting to the theorists of random functions. The two choices being of the same nature, let us examine the first. To randomize a category of signals is not easy work, because, apart from the filtered Gaussian white noise, the signals are not quite defined by their correlation functions of the first orders, in the same manner that a distribution function (density of probability), when multimodal, is not quite defined by its first moments.

For instance, a process of Poisson, with similar pulses appearing at random intervals of time, is not clearly described by considerations of correlation and spectrum. The proof of this inadequacy is simply that if a signaling series of pulses is mixed with a noisy series of pulses of double height, the two series having the same law of probability (and being independant of each other), it is very easy to filter by a detection of the amplitudes of peaks. That seems paradoxical, since the two series have the same spectrum, the energy of the noisy one being four times the energy of the signaling series.

The Correlation Detector

The first aim of a correlation detector is to decide if, in a corrupted signal, a sinusoidal waveform is present.

If the frequency of the hypothetical waveform were known, it would be sufficient to use a classical band-pass, the transmission band of which is centered on the said frequency f_0. The cutoff frequencies of the filter should be separated by a Δf approximately equal to the reciprocal of the duration of the test. The indicial response of such a filter is a sinusoidal waveform, modulated by a low frequency signal, the duration of which is approximately the duration of the signal to be tested. The output of the filter is a modulated sinusoidal waveform, and the probability of a sinusoidal component in the tested signal is related to the energy of the output. Such a filter is very fficult to design, because it has to store energy during ong time, and this is possible only with a great number lumped elements. The correlation method is a by-pass hich avoids this difficulty; as we are not interested in generating the waveform itself, but in detecting its esence or absence, we have to integrate the energy of e output of the filter. It appears that the combination the two operations of filtering and integrating can be ne more economically if we substitute for them equivent operations, for instance two modulations and two adratic integrations. It is easy to show that this new mbination is equivalent to the computation of an ement of the spectrum of the correlation function; narwing of Δf is automatically realized by the broadening of servation time, at the end of which the total filtered ergy is stored to be compared to a reference level.[1]

When the frequency of the sinusoidal waveform is known, the elementary solution of the problem is early to apply the corrupted signal at the input of a llection of band-pass filters with adjacent transmission ands, and to detect the largest output.

Of course, the comparison can be made only if the ansmission bands are related to the density spectrum of e supposed noise. The filtering must be as sharp as the uration of the signal allows. For long durations, we all have an unthinkable number of filters. The correlaon method permits avoiding this difficulty, but without hausting the information. This would require that the rrelation function should be analyzed in its Fourier mponents, which is the same problem as that first oposed. However, one step has been made; instead of e spectrum of the corrupted signal, we have to deal ith its square, so that the peaks are more apparent.[2,3]

The Limiting of Noise Peaks—Peculiar Case of Compression-Expansion

It can be shown that a peak limiting can improve the gnal/noise ratio. This can advantageously happen nly where: 1) The noise peaks are significantly high; and) The signal and the noise *are not* both Gaussian.

When the noise and the signal are almost Gaussian, limitation of peaks, or some other nonlinear detection, does not improve the classical results very much (with linear filtering).

When the noise, being not Gaussian, can be defined by an instantaneous transformation of a Gaussian noise, we can, by using an expansion or compression of amplitudes, put ourselves in conditions where the linear filtering is adequate; the transformation we apply to the corrupted signal does not affect the results obtainable by the method of maximum likelihood, which are invariable under every one-to-one change of variables.

The Maximum Likelihood Method

The maximum likelihood method, consisting in considering as the best regenerated signal the signal for which the probability density function of the corrupted signal is the greatest, is very easy to explain theoretically. It is more difficult to compute the error made in the application, because the classical methods of computation give the error only in asymptotic form, *i.e.*, for a great number of observations.

Dealing with the technical realization, we have first to describe the signal with a certain number of parameters, to find (mathematically) the expression of the estimation of these parameters as a function of the observations made on the corrupted signal, and *then* to form a device calculating these estimated values as a function of the observations.

Of course, the regeneration of the signal by a generator having these estimated parameters as inputs is generally omitted as useless, as is the case for instance when we deal with target detection. But the computation (of course automatic) of the estimated parameters is very intricate when the distribution of noise and description of signal are not simple.

Conclusion

If we refer to the three divisions of research which are pointed out in the introduction, it is to be noted that the stochastic definition of noise and signal is most important in a filtering design for a particular aim, and all energies must be devoted to this task.

At the present time, these definitions exist but in such terms that they do not correspond to a single physical realization, except in the case of a filtered white Gaussian noise, so that the present practical methods lead to practical results only for linear filters and compressor-expandor arrangements or to theoretically equivalent arrangements, such as those using correlation methods.

[1] E. Reich and P. Swerling, "The detection of a sine wave in aussian noise," *J. Appl. Phys.*, vol. 24, pp. 289–296; March, 1953.

[2] M. Horowitz and A. A. Johnson, "Theory of noise in a corretion detector," IRE Trans., vol. IT-1, pp. 3–5; December, 1955.

[3] M. D. Indjoudjian, "Le filtrage et la prediction des messages elon Norbert Wiener," ("La Cybernétique, Théorie du Signal et e l'Information"), *Rev. d'Optique,* (Paris), pp. 35–53; 1951.

On a Cross-Correlation Property for Stationary Random Processes*

JOHN L. BROWN, JR.†

Summary—Given two stationary random processes $x_1(t)$ and $x_2(t)$, the cross-correlation property of interest is the following: If one of the two processes is distorted by an instantaneous nonlinear device, then the cross correlation after the distortion is proportional to the cross-correlation function prior to the distortion.

Using an expansion of the second-order joint probability distribution $p(x_1, x_2)$ introduced by Barrett and Lampard, a necessary and sufficient condition for the above cross-correlation property is given in terms of requirements on the expansion coefficients.

In certain cases, the constant of proportionality involved in the cross-correlation property is equal to the "equivalent gain" of the nonlinear device as defined by Booton. A necessary and sufficient condition for these two constants to be identical is formulated in terms of the expansion coefficients of $p(x_1, x_2)$. The class of distributions satisfying this condition is a subclass of the set of distributions for which the cross-correlation property is valid.

Introduction

THE cross-correlation property to be studied in this paper may be stated as follows: A: If $x_1(t)$ and $x_2(t)$ are a pair of stationary time series, one of which undergoes amplitude distortion in a fixed "instantaneous" nonlinear device, then the cross-correlation function after the distortion is proportional to the cross-correlation function before the distortion. The constant of proportionality depends on the particular nonlinear device considered, but is independent of time in the stationary case.

It was first shown by Bussgang[1] that property A holds when the joint probability distribution of $x_1(t)$ and $x_2(t)$, $p(x_1, x_2)$, is Gaussian. In attempting to generalize Bussgang's result to a wider class of distributions, Barrett and Lampard[2] expanded the second-order distribution $p(x_1, x_2)$ in a double series of orthonormal polynomials, the particular polynomials being determined by the first-order distributions, $p_1(x_1)$ and $p_2(x_2)$. When the coefficient matrix of this expansion is a diagonal matrix, they demonstrated that property A is valid even in the more general case when x_1 and x_2 are nonstationary.

The class, Λ, of second-order distributions having a diagonal coefficient matrix with respect to the expansion includes the Gaussian distribution (thus giving Bussgang's result as a special case) and several other distributions of practical importance.[3]

In this paper, we shall make use of the Barrett-Lampard expansion for $p(x_1, x_2)$ and derive a necessary and sufficient condition for property A in terms of the expansion coefficients. We shall also discuss the relation between the constant of proportionality involved in the cross-correlation property and Booton's "equivalent gain"[4] for the nonlinear device under consideration, again deriving a necessary and sufficient condition for the two quantities to be identical.

General Properties of the Expansion

We let $p(x_1, x_2)$ denote the second-order joint probability distribution[5] of x_1 and x_2, noting that this distribution depends on $\tau = t_2 - t_1$ when x_1 and x_2 are stationary processes. Two sets of orthonormal polynomials $\{\theta_n^{(1)}(x_1)\}$ and $\{\theta_n^{(2)}(x_2)\}$ are constructed having weighting functions $p_1(x_1)$ and $p_2(x_2)$ respectively, where $p_1(x_1)$ and $p_2(x_2)$ are the corresponding first-order probability distributions. These first-order distributions are given by[6]

$$\left.\begin{aligned} p_1(x_1) &= \int p(x_1, x_2)\, dx_2 \\ p_2(x_2) &= \int p(x_1, x_2)\, dx_1 \end{aligned}\right\}. \tag{1}$$

The expansion of $p(x_1, x_2)$ in terms of the polynomials has the form

$$p(x_1, x_2) = p_1(x_1)p_2(x_2) \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn}(\tau)\theta_m^{(1)}(x_1)\theta_n^{(2)}(x_2) \tag{2}$$

where

$$a_{mn}(\tau) = a_{mn} = \iint p(x_1, x_2)\theta_m^{(1)}(x_1)\theta_n^{(2)}(x_2)\, dx_1\, dx_2 \tag{3}$$

and the orthonormality conditions are[7]

$$\left.\begin{aligned} \int p_1(x_1)\theta_m^{(1)}(x_1)\theta_n^{(1)}(x_1)\, dx_1 &= \delta_{mn} \\ \int p_2(x_2)\theta_m^{(2)}(x_2)\theta_n^{(2)}(x_2)\, dx_2 &= \delta_{mn} \end{aligned}\right\}. \tag{4}$$

Since $p(x_1, x_2)$, $p_1(x_1)$, and $p_2(x_2)$ are probability distributions, the following relations obtain

* Manuscript received by the PGIT, August 7, 1956.

† Ordnance Res. Lab., Pennsylvania State University, Pa.

[1] J. J. Bussgang, "Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals," Mass. Inst. Tech., Res. Lab. Electronics, Cambridge, Mass., Tech. Rep. No. 216; March 26, 1952.

[2] J. F. Barrett and D. G. Lampard, "An expansion for some second-order probability distributions and its application to noise problems," IRE Trans., vol. IT-1, pp. 10–15; March, 1955.

[3] *Ibid.*

[4] R. C. Booton, Jr., "The analysis of nonlinear control systems with random inputs," *Proc. Symposium on Nonlinear Circuit Analysis*, Bklyn. Polytechnic Inst., pp. 369–391; April 23–24, 1953.

[5] That is, $p(x_1, x_2)dx_1\, dx_2$ gives the joint probability of the inequalities $x_1 \leq x_1(t_1) \leq x_1 + dx_1$ and $x_2 \leq x_2(t_2) \leq x_2 + dx_2$. This function is commonly referred to as the "joint frequency" or "joint density" function in the literature.

[6] All integrals are to be understood as extending over the whole range of variation of the random variables involved.

[7] δ_{mn} is the usual Kronecker delta defined by $\delta_{mn} = \begin{cases} 1 \text{ if } m = n \\ 0 \text{ if } m \neq n \end{cases}$

$$\left.\begin{aligned}
&\iint p(x_1, x_2)\,dx_1\,dx_2 = 1\\
&\int p_1(x_1)\cdot 1\cdot 1\,dx_1 = \int p_2(x_2)\cdot 1\cdot 1\,dx_2 = 1\\
&\int p_1(x_1)(x_1 - \mu_1)\cdot 1\,dx_1 = 0\\
&\int p_2(x_2)(x_2 - \mu_2)\cdot 1\,dx_2 = 0\\
&\int p_1(x_1)\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_1 - \mu_1}{\sigma_1}\right)dx_1 = 1\\
&\int p_2(x_2)\left(\frac{x_2 - \mu_2}{\sigma_2}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right)dx_2 = 1
\end{aligned}\right\} \quad (5)$$

ıere μ_1 and σ_1 are respectively the mean and standard viation of $x_1(t)$, and μ_2, σ_2 are the corresponding quantis for $x_2(t)$.

These relations imply

$$\left.\begin{aligned}
\theta_0^{(1)}(x_1) &= \theta_0^{(2)}(x_2) \equiv 1\\
\theta_1^{(1)}(x_1) &= \frac{x_1 - \mu_1}{\sigma_1}\\
\theta_1^{(2)}(x_2) &= \frac{x_2 - \mu_2}{\sigma_2}
\end{aligned}\right\}. \quad (6)$$

Now consider $\langle[\theta_m^{(1)}(x_1) + \lambda\theta_n^{(2)}(x_2)]^2\rangle$.[8]

This average is equal to

$$\iint p(x_1, x_2)[\theta_m^{(1)}(x_1) + \lambda\theta_n^{(2)}(x_2)]^2\,dx_1\,dx_2,$$

hich reduces to the quantity

$$1 + 2\lambda a_{mn} + \lambda^2,$$

ing expansion (2) for $p(x_1, x_2)$ and the orthonormal lations (4). Since this quantity must be non-negative r all real values of λ, it is easily verified that

$$|\,a_{mn}\,| \leq 1 \quad \text{for all } m,n. \quad (7)$$

In particular, $a_{nn}^2 \leq 1$, as established by Barrett and ampard in a similar manner for the special case in which $a_{mn}\}$ is a diagonal matrix.

If we define the cross correlation (un-normalized) etween $x_1(t)$ and $x_2(t)$ by

$$_{12}(\tau) = \iint (x_1 - \mu_1)(x_2 - \mu_2)p(x_1, x_2)\,dx_1\,dx_2$$
$$= \langle(x_1 - \mu_1)(x_2 - \mu_2)\rangle, \quad (8)$$

hen, by using the expansion for $p(x_1, x_2)$ and the orthoormal properties of the polynomials, it can be shown that

$$a_{11}(\tau) = \frac{\langle(x_1 - \mu_1)(x_2 - \mu_2)\rangle}{\sigma_1\sigma_2} = \frac{\psi_{12}(\tau)}{\sigma_1\sigma_2}. \quad (9)$$

Another important property of the coefficients a_{mn} in (2) is that $a_{n0} = a_{0n} = 0$ for $n \neq 0$. This may be shown as follows

$$p_2(x_2) = \int p(x_1, x_2)\,dx_1$$
$$= \int p_1(x_1)p_2(x_2)\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{mn}\theta_m^{(1)}(x_1)\,\theta_n^{(2)}(x_2)\,dx_1$$
$$= \sum_{n=0}^{\infty} a_{0n}\theta_n^{(2)}(x_2)p_2(x_2)$$

upon interchanging the operations and using the orthonormality relations (4) to simplify the expression. For $p_2(x_2) \not\equiv 0$, we conclude that

$$\sum_{n=0}^{\infty} a_{0n}\theta_n^{(2)}(x_2) = 1.$$

But, this equality indicates that the quantity on the left is the expansion of the function which is identically equal to "one" over the range of variation of x_2. Since $\theta_0^{(1)}(x_1) = 1$ from (6), and since the expansion is assumed to exist and to be unique,

$$a_{00} = 1 \quad \text{and} \quad a_{0n} = 0 \quad \text{for} \quad n > 0. \quad (10)$$

A similar argument shows $a_{n0} = 0$ for $n > 0$.

Analysis[9]

The cross-correlation property A is expressed mathematically by

$$A: \qquad \Psi_{12}(\tau) = K(f)\;\psi_{12}(\tau), \quad (11)$$

where

$$\Psi_{12}(\tau) = \iint [f(x_1) - \langle f(x_1)\rangle](x_2 - \mu_2)p(x_1, x_2)\,dx_1\,dx_2, \quad (12)$$

and

$$\psi_{12}(\tau) = \iint (x_1 - \mu_1)(x_2 - \mu_2)p(x_1, x_2)\,dx_1\,dx_2. \quad (13)$$

$K(f)$ is a linear functional depending on the function $f(x)$, which characterizes the nonlinear device and is independent of τ.

Actually, the term $\langle f(x_1)\rangle$ in the integrand of (12) may be taken as zero, since this term contributes nothing to the integral as seen from the following argument:

$$\iint \langle f(x_1)\rangle\,(x_2 - \mu_2)p(x_1, x_2)\,dx_1\,dx_2$$
$$= \langle f(x_1)\rangle \iint (x_2 - \mu_2)p(x_1, x_2)\,dx_1\,dx_2$$
$$= \langle f(x_1)\rangle \int p_2(x_2)(x_2 - \mu_2)\,dx_2$$
$$= \langle f(x_1)\rangle\,(\mu_2 - \mu_2) = 0.$$

[8] "⟨ ⟩" denotes ensemble averaging.

[9] The arguments and manipulations used in this paper are purely formal. No questions of convergence or legitimacy of interchanging operations are considered here. It is assumed that the class of functions treated is suitably restricted so that the required expansions exist.

Thus, $\Psi_{12}(\tau)$ may be written equivalently as

$$\Psi_{12}(\tau) = \iint f(x_1)(x_2 - \mu_2)p(x_1, x_2)\,dx_1\,dx_2. \tag{12a}$$

Theorem 1

For stationary processes, $\Psi_{12}(\tau) = K(f)\psi_{12}(\tau)$, if and only if there exists a sequence of real constants $\{d_m\}$, $m = 1, 2, \cdots$ independent of τ, with $d_1 = 1$ such that $a_{m1}(\tau) = d_m a_{11}(\tau)$ for $m = 1, 2, \cdots$.

Proof

We first show that the stated condition implies property A. Let

$$f(x_1) = \sum_{k=0}^{\infty} c_k \theta_k^{(1)}(x_1),$$

with

$$c_k = \int f(x_1)\theta_k^{(1)}(x_1)p_1(x_1)\,dx_1.$$

Then,

$$\langle f(x_1)\rangle = \int f(x_1)p(x_1)\,dx_1 = c_0.$$

Thus,

$$f(x_1) - \langle f(x_1)\rangle = \sum_{k=1}^{\infty} c_k\theta_k^{(1)}(x_1)$$

and

$$\Psi_{12}(\tau) = \iint\left[\sum_{k=1}^{\infty} c_k\theta_k^{(1)}(x_1)\right]\cdot(x_2 - \mu_2)$$
$$\cdot\left[\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{mn}\theta_m^{(1)}(x_1)\theta_n^{(2)}(x_2)\right]\cdot p_1(x_1)p_2(x_2)\,dx_1\,dx_2.$$

But

$$x_2 - \mu_2 = \sigma_2\theta_1^{(2)}(x_2).$$

Using relations (4) and interchanging operations, the expression for $\Psi_{12}(\tau)$ reduces to

$$\Psi_{12}(\tau) = \sigma_2\sum_{k=1}^{\infty} a_{k1}(\tau)c_k. \tag{14}$$

By assumption, $a_{k1} = d_k a_{11}$ with $d_1 = 1$. Substituting in (14)

$$\Psi_{12}(\tau) = \left[\sigma_2\sum_{k=1}^{\infty} d_k c_k\right]a_{11}(\tau).$$

From (9)

$$\Psi_{12}(\tau) = \left(\sum_{k=1}^{\infty}\frac{d_k c_k}{\sigma_1}\right)\psi_{12}(\tau). \tag{15}$$

Thus, defining $K(f)$ as $\sum_{k=1}^{\infty} d_k c_k/\sigma_1$,[10] (15) becomes

$$\Psi_{12}(\tau) = K(f)\psi_{12}(\tau) \qquad \text{as required.}$$

Conversely, assume that for each $f(x)$, a real number $K(f)$ is given such that property A holds. Substituting the expansion for $p(x_1, x_2)$ in the integrands of both $\Psi_{12}(\tau)$ and $\psi_{12}(\tau)$ and reducing, results in the expression

$$\sum_{k=1}^{\infty} a_{k1}(\tau)c_k = K(f)a_{11}(\tau)\sigma_1. \tag{16}$$

Since $K(f)$ is assumed to be defined for all $f(x)$, let $K[\theta_m^{(1)}(x_1)] = h_m$. The $\{c_k\}$ corresponding to $f(x) = \theta_m^{(1)}(x_1)$ are $c_k = \delta_{km}$. For this choice of $f(x)$, (16) gives

$$a_{k1} = h_m a_{11}\sigma_1 \quad \text{for} \quad k = 1, 2, \cdots. \tag{17}$$

Note in the above that $h_1 = 1/\sigma_1$. Defining $d_m = h_m\sigma_1$, we have from (17),

$$a_{m1} = d_m a_{11} \quad \text{for} \quad m = 1, 2, \cdots \quad \text{with} \quad d_1 = 1$$

as required.

In the case where $p(x_1, x_2)$ has a diagonal matrix $a_{m1} = 0$ for $m \neq 1$, and, consequently, we may choose the trivial sequence $\{d_1 = 1, d_n = 0 \text{ for } n > 1\}$ to show that the cross-correlation property holds. This special result for diagonal matrices was previously established by Barrett and Lampard in their paper.

Relation to Booton's "Equivalent Gain"

In his analysis[11] Booton approximates the output $f(x_1)$, of an instantaneous nonlinear device by

$$f(x_1) \doteq K_b(f)(x_1 - \mu_1) \tag{18}$$

where $x_1(t)$ is the input, and $K_b(f)$ is a constant to be determined. The criterion used by Booton to fix $K_b(f)$ is that the mean square difference between $K_b(f)\cdot(x_1 - \mu_1)$ and $f(x_1)$ be made a minimum; that is,

$$\int [f(x_1) - K_b(f)\cdot(x_1 - \mu_1)]^2 p_1(x_1)\,dx_1$$

is to be minimized. Expanding and setting the derivative with respect to K_b equal to zero gives the formula

$$K_b(f) = \int f(x_1)\frac{(x_1 - \mu_1)}{\sigma_1^2}p_1(x_1)\,dx_1. \tag{19}$$

$K_b(f)$ is then termed the "equivalent gain" of the nonlinear device. It is interesting to note that this is the same as the expression for the $K(f)$ appearing in the cross-correlation property as found by Barrett and Lampard.[12,13] Thus, for distributions having a diagonal coefficient matrix $\{a_{mn}\}$, the constant of proportionality in the cross-correlation property is identical to the "equivalent gain" of the nonlinear device being considered.

[10] This definition involves only the coefficients of $f(x_1)$ and the $\{d_k\}$, which are assumed to be known.

[11] Booton, *loc. cit.*

[12] Barrett and Lampard, *loc. cit.*

[13] In (34) A of the Barrett-Lampard paper, the factor $p_2(x_2)$ has been omitted from the integrand in the representation of the proportionality constant.

is natural to ask if this identity can be extended to a der class of distributions; the following theorem gives answer in terms of the expansion coefficients a_{mn}. We st note that (19) can be written

$$K_b(f) = \frac{c_1}{\sigma_1}, \tag{20}$$

ere $c_1 = \int f(x_1)\theta_1^{(1)}(x_1)p_1(x_1)dx_1$ as before. Since this always the representation for $K_b(f)$, the question reduces a matter of when $\Psi_{12}(\tau)$ is equal to $c_1/\sigma_1\ \psi_{12}(\tau)$.

eorem 2

$\Psi_{12}(\tau) = K_b\psi_{12}(\tau)$ if and only if $a_{m1}(\tau) = 0$ for $m \geq 2$.

oof

As outlined in the preceding theorem,

$$\Psi_{12}(\tau) = \sigma_2 \sum_{k=1}^{\infty} a_{k1}(\tau)c_k$$

d

$$\psi_{12}(\tau) = \sigma_1\sigma_2 a_{11}(\tau).$$

herefore, if $\Psi_{12}(\tau) = K_b\psi_{12}(\tau) = c_1/\sigma_1\ \psi_{12}(\tau)$, we have

$$\sigma_2 \sum_{k=1}^{\infty} a_{k1}c_k = c_1\sigma_2 a_{11} \quad \text{for all} \quad f(x_1),$$

$$\sum_{k=2}^{\infty} a_{k1}(\tau)c_k = 0 \quad \text{for all} \quad f(x_1). \tag{21}$$

The particular choice of $f(x_1) = \theta_m^{(1)}(x_1)$ yields $c_k = $ $_m$ and implies

$$a_{m1}(\tau) = 0 \text{ for } m \geq 2,$$

us establishing the necessity of the condition.

Conversely, if $a_{m1}(\tau) = 0$ for $m \geq 2$, then

$$\Psi_{12}(\tau) = \sigma_2 \sum_{k=1}^{\infty} a_{k1}(\tau)c_k = \sigma_2 a_{11}(\tau)c_1.$$

ut

$$a_{11}(\tau) = \frac{\psi_{12}(\tau)}{\sigma_1\sigma_2}$$

nd, therefore,

$$\Psi_{12}(\tau) = \frac{c_1}{\sigma_1}\psi_{12}(\tau)$$

$$= K_b\psi_{12}(\tau), \quad \text{as required.}$$

If we define Λ^* to be the class of distributions $p(x_1, x_2)$ r which $a_{m1} = 0$ $(m \geq 2)$, then the class of distributions ith diagonal matrices Λ[14] is included in Λ^*. Whenever (x_1, x_2) belongs to Λ^*, the constant of proportionality in he cross-correlation property is equal to the "equivalent gain" for the nonlinear device; further, $p(x_1, x_2)$ must belong to Λ^* for this to be true.

Generalization

The restriction to stationary processes in the above arguments is purely a matter of notational convenience, since only ensemble averages are involved, and time is regarded essentially as a parameter. In the event that $x_1(t)$ and $x_2(t)$ are nonstationary, a_{mn} depends on two parameters, t_1 and t_2, and both Ψ_{12} and ψ_{12} also depend on t_1 and t_2. The polynomials $\theta_m^{(1)}(x_1)$ and $\theta_n^{(2)}(x_2)$ will, in general, depend on t_1 and t_2 respectively, and, consequently, the expansion coefficients of $f(x_1)$ will be functions of t_1. $K(f)$ will no longer be independent of time, but will likewise depend on the parameter, t_1; however, the constants d_m of theorem 1 are still required to be real constants exhibiting no time dependence.

Conclusion

The correlation property treated in this paper states that the cross correlation between two random processes, one of which has undergone an instantaneous nonlinear distortion, is proportional to the cross correlation before distortion. A necessary and sufficient condition for this property to hold has been given in terms of the coefficient matrix obtained when, $p(x_1, x_2; \tau)$ is expanded in a double series of orthonormal polynomials determined by the first-order probability distributions of $x_1(t)$ and $x_2(t)$. This condition is essentially equivalent to requiring that the property hold for the infinite sequence of nonlinear devices, $\{F_n\}$, where the response of F_n to an input x is $\theta_n^{(1)}(x)$. Thus, if the property holds for this sequence of "orthonormal nonlinear devices," it will hold for an arbitrary instantaneous nonlinear device and conversely.

We have also determined under what conditions the constant of proportionality involved in the correlation property is equal to Booton's "equivalent gain" for the nonlinear device. The condition stated shows that whenever the two constants are equal, the cross-correlation property also holds, although the converse is not necessarily true.

For Gaussian processes, Booton's analysis shows that the difference $f(x_1) - K_bx_1$ is uncorrelated with x_1; that is, the output of the nonlinear device can be represented as the sum of two terms, where one of the terms is the result of a linear operation on the input, and the other is uncorrelated with the input. If the correlation property holds, then this type of representation for the output of the nonlinear device is always possible. Thus, in any case where the random process considered is such that $p(x_1, x_2)$ belongs to Λ^*, Booton's approximation of $f(x_1)$ by K_bx_1 results in an error term which is uncorrelated with the input, x_1. Since the second-order distribution of a constant amplitude sine wave belongs to Λ^* as shown by Barrett and Lampard, the linearization effected by the "describing function" is also of the type leading to an error which is uncorrelated with the sine wave input.

[14] *Ibid.*

A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena —Part I*

D. A. DARLING† AND A. J. F. SIEGERT‡

Summary—The problem of finding the probability of distribution of the functional

$$\int_{t_0}^{t} \Phi(X(\tau), \tau)\, d\tau,$$

where $X(\tau)$ is a (multidimensional) Markoff process and $\Phi(X, \tau)$ is a given function, appears in many forms in the theory of noise and other random phenomena. We have shown that a certain function from which this probability distribution can be obtained is the unique solution of two integral equations. We also developed a perturbation formalism which relates the solutions of the integral equations belonging to two different functions $\Phi(X, \tau)$. If the transition probability density for $X(\tau)$ is the principal solution of two partial differential equations of the Fokker-Planck-Kolmogoroff type, the principal solution of two similar differential equations is the solution of the integral equations. As an example, we calculated the probability distribution of the sample probability density for a stationary Markoff process.

Introduction

IN THE theory of noise and similar random phenomena, a small number of special problems have been solved by various special methods. Each of these methods seems to apply only to the particular problem for which it was developed or at best to a rather restricted class of problems. It seemed of interest, therefore, to develop a systematic approach to a wider class of problems, which contains as special cases most of the problems solved before. Even though this approach leads to rather formidable integral or differential equations, so that the number of new problems which can be solved exactly will be small, it leads to a perturbation formalism for problems "in the neighborhood" of those permitting exact solutions.

We consider the problem of finding the probability distribution of the random variable

$$u = \int_{t_0}^{t} \Phi(X(\tau), \tau)\, d\tau \tag{1}$$

where $X(\tau)$ is a Markoff process with components $x_1(\tau)$, $x_2(\tau) \cdots x_n(\tau)$. This problem arose originally as the problem of finding the probability distribution of the noise output of a radio receiver consisting of a linear amplifier, an arbitrary detector, and a second linear amplifier. Let $x(\tau)$ be the output voltage of the first amplifier at a time $\tau \geq 0$ before observation,[1] $\varphi[x(\tau)]$ the output voltage of the detector at the same time, and $K(\tau)$ the output of the second amplifier at the time of observation if a δ-function pulse is applied to it at the time τ. The output voltage V of the second amplifier in response to $x(\tau)$ is then

$$V = \int_0^t K(\tau)\varphi(x(\tau))\, d\tau \tag{2}$$

if the noise was turned on at time $t \geq 0$ before observation. If the input of the first amplifier is white noise, $x(\tau)$ is a Gaussian random function. This fact has made it possible to reduce the problem for the special case $\varphi(x) \equiv x^2$ to the solution of an integral equation in one variable only. Except for this special form of the detector function, and of course in the trivial case $\varphi(x) \equiv x$, the Gaussian property of $X(\tau)$ does not simplify the problem.

If the first amplifier is equivalent to a network with lumped circuit elements and its input is white noise, $x(\tau)$ is also a component of a Markoff process. This led us to consider the more general problem stated above which has many applications apart from the noise output of radio receivers.

If, for instance, a domain $\mathfrak{D}$ is chosen in X space, and $\Phi(X, \tau)$ is defined by

$$\Phi(X, \tau) = \begin{cases} 1 & \text{when } X \text{ is in } \mathfrak{D} \\ 0 & \text{otherwise} \end{cases}$$

then u is that part of the time $(t - t_0)$ during which X is in $\mathfrak{D}$ in the time interval $(t - t_0)$, and $(t - t_0)^{-1}u$ is an estimate for the probability that X is in $\mathfrak{D}$, obtained from the finite sample. The distribution of u is thus of importance if it is desired to estimate the accuracy with which the probability distribution of time homogeneous processes $X(\tau)$ can be obtained from finite samples.

If, specially, $\mathfrak{D}$ is defined by

$$x_1 < a,$$

with $x_2, x_3 \cdots x_n$ unrestricted, the probability for $u = t - t_0$ is the probability that $x_1(t') < a$ for all t' in $t_0 < t' < t$ (except for a set of measure zero) and, for a continuous function $x_1(\tau)$, this is the cumulative distribution of the absolute maximum of $x_1(t)$ in the interval

* Manuscript received by the PGIT, August 13, 1956. Most of this work was done while the authors were consultants for The RAND Corp., Santa Monica, Calif.

† Dept. of Mathematics, University of Michigan, Ann Arbor, Mich.

‡ Dept. of Physics, Northwestern University, Evanston, Ill.

[1] It is convenient to choose the time scale positive into the past.

t), if it is considered as function of a, or the cumulative bability distribution of the one-sided first-passage ie, if it is considered as function of t.

f $\mathfrak{D}$ is defined by

$$b < x_1 < a$$

h $x_2, x_3, \cdots x_n$ unrestricted, one obtains in a similar v for continuous $x_1(\tau)$ the distribution of the two-sided t-passage time (escape time), and the distribution of range $r \equiv \max_\tau x(\tau) - \min_\tau x(\tau)$, $t_0 < \tau < t$. For -dimensional Markoff processes the problem of first-ssage time, range, and maximum has been solved by an ler method.[2]

The problem of finding the distribution of the empirical ectrum or the Fourier coefficients obtained from a mple can also be formulated as a special case of our oblem. If $\psi_1(\tau), \psi_2(\tau), \cdots \psi_l(\tau)$ are given functions, instance trigonometric functions, the characteristic nction

$$\left\langle \exp\left\{ i \sum_{k=1}^{l} \zeta_k \int_{t_0}^{t} x_1(\tau)\psi_k(\tau)\, dt \right\}\right\rangle_{\text{av}}$$

the joint distribution of the Fourier coefficients is also e characteristic function

$$\left\langle \exp\left\{ i\zeta \int_{t_0}^{t} \Phi(x_1(\tau), \tau)\, d\tau \right\}\right\rangle_{\text{av}}$$

the random variable u defined by (1), if one chooses

$$\Phi(X(\tau), \tau) \equiv \zeta^{-1} x_1(\tau) \sum_{k=1}^{l} \zeta_k \psi_k(\tau). \tag{3}$$

TEGRAL EQUATIONS FOR THE CONDITIONAL CHARACTERISTIC FUNCTION, PERTURBATION FORMULA, AND DIFFERENTIAL EQUATION

In the present paper we present a heuristic derivation[3] two integral equations for the function

$$X_0, t_0 \mid X, t; \lambda) \equiv \left\langle \exp\left\{ -\lambda \int_{t_0}^{t} \Phi(X(\tau), \tau)\, d\tau \right\} \;\middle|\; X(t_0) = X_0, X(t) = X \right\rangle_{\text{av}} \cdot p(X_0, t_0 \mid X, t) \tag{4}$$

here $\langle | \rangle_{av}$ denotes the average of the functional on the ft of the vertical bar under the condition written on its ght, and where $p(X_0, t_0) \mid X, t)dX$ is the probability that (t) is in the volume element dX at X, if $X(t_0) = X_0$. The rameter λ will be chosen positive real if Φ is non-negative, and negative imaginary otherwise. The Fourier or Laplace transform of the probability density for the variable u defined by (1) is obviously r/p if initial and end conditions are imposed and $\int r dX$ if only initial conditions are imposed, and $\int W(X_0, t_0) r dX dX_0$ if no conditions are imposed, where $W(X_0, t_0)$ is the probability density for $X(t_0)$.

Consider now $X(\tau)$ as the path of a particle in X space. If, at first, $\Phi(X(\tau), \tau)$ is assumed to be non-negative and λ real and positive, then $\lambda\Phi(X, \tau)d\tau$ can be interpreted as the probability of a "collision" at the point X in the time interval $(\tau, \tau + d\tau)$. A "collision" is thereby understood to be an event which does not affect the path of the particle nor the probability of later collisions, but leaves a mark on the particle so that the number of collisions experienced by the particle can be counted. The functional $\exp\,[-\lambda \int_{t_0}^{t} \Phi(X(\tau), \tau)d\tau]$ is thus the probability that the particle suffers no collisions on a path $X(\tau)$ which leads from X_0 to X; and $r(X_0, t_0 \mid X, t; \lambda)\, dX$ is the probability of finding the particle at time t in the volume element dX at X without any marks, if it started at X_0 at time t_0. An integral equation for $r(X_0 t_0 \mid X, t; \lambda)$ is obtained by subtracting from $p(X_0, t_0 \mid X, t)\, dX$ the probability that the particle reached some point X' in the volume element dX' at some time t' without collisions, suffered the first collision there in the time interval $(t', t' + dt')$, and went on from there to X suffering an arbitrary and irrelevant number of collisions. One thus has the integral equation

$$r(X_0, t_0 \mid X, t; \lambda) = p(X_0, t_0 \mid X, t) - \lambda \int_{t_0}^{t} dt' \int dX' \cdot r(X_0, t_0 \mid X', t'; \lambda)\Phi(X', t')p(X', t' \mid X, t). \tag{5}$$

Repeating the same argument with the last collision one obtains

$$r(X_0, t_0 \mid X, t; \lambda) = p(X_0, t_0 \mid X, t) - \lambda \int_{t_0}^{t} dt' \int dX' \cdot p(X_0, t_0 \mid X', t')\Phi(X', t')r(X', t' \mid X, t; \lambda). \tag{6}$$

A formal derivation which removes the restrictions $\lambda > 0$, $\Phi(X(\tau), \tau) \geq 0$ is given in Appendix I.

Since the integral equations will in general be difficult to solve in closed form, a perturbation formalism seems of value. Suppose that the solution of the integral equations is $r_1(X_0, t_0 \mid X, t; \lambda)$ for $\Phi(X, t) \equiv \Phi_1(X, t)$. Let now a second scattering medium be added, such that the probability of a collision at X in $(t, t + dt)$ is increased by $\lambda[\Phi_2(X, t) - \Phi_1(X, t)]dt$. By repetition of the argument given above one then obtains two integral equations for the solution $r_2(X_0, t_0 \mid X, t; \lambda)$ of (5) and (6), when $\Phi(X, t) \equiv \Phi_2(X, t)$:

$$r_2(X_0, t_0 \mid X, t; \lambda) = r_1(X_0, t_0 \mid X, t; \lambda) - \lambda \int_{t_0}^{t} dt' \int dX' r_2(X_0, t_0 \mid X', t') \cdot [\Phi_2(X', t') - \Phi_1(X', t')]r_1(X', t' \mid X, t; \lambda) \tag{7}$$

[2] A. J. F. Siegert, "On the first passage time probability problem," ys. Rev., vol. 81, pp. 617–623; February 15, 1951.
D. A. Darling and A. J. F. Siegert, "The first passage problem a continuous Markoff process," *Ann. Math. Stat.*, vol. 24, pp. 4–639; December, 1953, and work quoted there.

[3] A rigorous derivation has been given by D. A. Darling and A. F. Siegert, "On the Distribution of Certain Functionals of Markoff ocesses," The RAND Corp., Rep. P-429; April, 1954. Appeared abbreviated form, *Proc. Natl. Acad. Sci.*; August, 1956. A short etch of the method was given by Siegert, "Passage of stationary ocesses through linear and non-linear devices," IRE TRANS., vol. IT-3, pp. 4–25; March, 1954.

and

$$r_2(X_0, t_0 \mid X, t; \lambda) = r_1(X_0, t_0 \mid X, t; \lambda)$$
$$- \lambda \int_{t_0}^{t} dt' \int dX' r_1(X_0, t_0 \mid X', t')$$
$$\cdot [\Phi_2(X', t') - \Phi_1(X', t')] r_2(X', t' \mid X, t; \lambda). \quad (8)$$

We can thus obtain successive approximations for r_2 if r_1 is known, by the usual iteration procedure.

A formal derivation independent of the restrictions imposed on λ and Φ has been given.[3] This derivation serves, furthermore, to prove the uniqueness of the solutions of (5) and (6).

In many cases of practical interest $p(X_0, t_0 \mid X, t)$ is the principal solution of two partial differential equations of the form

$$\frac{\partial p}{\partial t} = Lp \quad (9)$$

$$-\frac{\partial p}{\partial t_0} = L_0^+ p \quad (10)$$

where L is defined by[4]

$$Lp \equiv \frac{1}{2} \sum_{kl} \frac{\partial^2}{\partial x_k \partial x_l} [B_{kl}(X, t)p] - \sum_k \frac{\partial}{\partial x_k} [A_k(X, t)p] \quad (11)$$

and L_0^+ is the adjoint of this operator with X and t replaced by X_0 and t_0. The physical meaning of (10) is that of a continuity equation

$$\frac{\partial p}{\partial t} = -\text{Div } J \quad (12)$$

where Div is the divergence and J is the probability current in X space. The current J can be interpreted as a diffusion current with components $-\frac{1}{2} \sum_l B_{kl}\, \partial p/\partial x_l$ and a drift current with components $(A_k - \frac{1}{2} \sum_l \partial B_{kl}/\partial x_l)p$ if one wants to keep the form $-\frac{1}{2} B \cdot \text{Grad } p$ for the diffusion current. (If one prefers to retain the drift current in the form $V_{av}p$, where V_{av} is the average velocity one interprets $A_k p$ as the drift current and $-\frac{1}{2} \sum_l \partial/\partial x_l\, (B_{kl}p)$ as diffusion current.) Formal application of the operator $L - \partial/\partial t$ to (5) yields

$$\left(L - \frac{\partial}{\partial t}\right) r = \lambda \int dX' r(X_0, t_0 \mid X', t; \lambda)$$
$$\cdot \Phi(X', t) p(X', t \mid Xt). \quad (13)$$

With the initial condition

$$p(X', t \mid X, t) = \delta(X' - X) \quad (14)$$

this becomes

$$\frac{\partial}{\partial t} r(X_0, t_0 \mid X, t; \lambda) = \{L - \lambda \Phi(X, t)\} r(X_0, t_0 \mid X, t; \lambda). \quad (15)$$

The interpretation of (15) is clearly that the rate o particle loss by collisions $\lambda\Phi(X, t)$ has been added to th continuity equation. Formal application of the operat $L_0^+ + \partial/\partial t_0$ to (6) yields in the same way the differenti equation

$$-\frac{\partial}{\partial t_0} r(X_0, t_0 \mid X, t, \lambda)$$
$$= \{L_0^+ - \lambda \Phi(X_0, t_0)\} r(X_0, t_0 \mid X, t; \lambda). \quad (16$$

We showed[3] that the principal solution of either (15) o (16), if it exists, is actually a solution of (5) and (6) an is, therefore, by virtue of the uniqueness theorem, th solution of these integral equations.

In special cases, (5), (6), (13), and (15) reduce to th integral and differential equations derived by Kac Rosenblatt,[6] and Fortet.[7] When $X(\tau)$ is taken to be th one-dimensional Wiener function $x(\tau)$ (once integrate "white noise") with $x(0) = 0$, and $\Phi[X(\tau), \tau] = V(x$ one obtains from (5) the integral equation (3.8) of Kac and the Laplace transform of (15) reduces to equatio (3.14) of Kac.[5] When the components $x_k(\tau)$ of $X(\tau)$ ar Wiener functions with $x_k(0) = 0$, (5) reduces to equatio (1.9) of Rosenblatt.[6] The differential equation (16) wa derived directly by Fortet.[7]

If $X(\tau)$ is a Gaussian Markoff process, and $\Phi(X(\tau), \tau) =$ $K(t)x_1^2(\tau)$ the method of Kac and Siegert[8] can be applie also, and leads to an integral equation in the time variabl only. In this case the solution of (15) is an exponentia function of a second-degree polynomial in the component of X_0 and X, and (15) leads to first-order nonlinea differential equations for the coefficients. The equivalenc of these to the integral equation of Kac and Siegert[8] requires a somewhat lengthy discussion and will be given i Part II of this paper. It seems interesting to note that th present procedure yields some of the results of Kac an Siegert[8] in closed form.

Example: The Distribution of the Sample Probability Density for a Stationary Markoff Process

It is often necessary to infer the probability densit for a random process from a sample. If the process i stationary a convenient estimate $w^*(z)$ of the probabilit density $w(z)$ is the fraction of the sample length durin which the value of the random process $X(\tau)$ lies in a smal interval or volume element Δ centered on z, divided by Δ

The calculation will be carried through for the Markof

[4] M. C. Wang and G. E. Uhlenbeck, "On the theory of the Brownian motion II," *Rev. Mod. Phys.*, vol. 17, pp. 323–342; April–July; 1945.
A. Kolmogoroff, "Uber die analytischen methoden in der wahrscheinlichkeitsrechrung," *Math. Ann.*, vol. 104, pp. 415–458; March, 1931.

[5] M. Kac, "On some connections between probability theory an differential and integral equations," *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability*, University o California Press, Berkeley, Calif., pp. 189–215; 1951.

[6] M. Rosenblatt, "On a class of Markov processes," *Trans. Amer Math. Soc.*, vol. 71, pp. 120–135; July, 1951.

[7] A. Blanc-Lapierre and R. Fortet, "Théorie des Fonction Aléatoires," Masson et C^{ie}, Paris, 321 pp.; 1953.

[8] M. Kac and A. J. F. Siegert, "On the theory of noise in radi receivers with square law detectors," *J. Appl. Phys.*, vol. 18, pp 383–397; April, 1947. For an improvement of this method, se Siegert, "Passage of stationary processes through linear and nonlinear devices," IRE Trans., vol. PGIT-3, pp. 4–25; March, 1954.

cess with one component $x(\tau)$; the generalization to n-dimensional process is trivial, if one is interested in joint distribution of the components. [It must be phasized, however, that the generalization to the ple probability of *one* component of a *multidimensional* rkoff process is not trivial, since the integral equation does not simplify appreciably in that case.] It would easy, on the other hand, to generalize our calculation obtain the joint distribution of $w^*(z_1)$, $w^*(z_2)$ $\cdots$ $z_k)$. This distribution may be useful in obtaining an roximation to the distribution of

$$\int_0^t \Phi(x(\tau))\,d\tau \equiv \int \Phi(z)w^*(z)\,dz,$$

(z) and $w^*(z)$ are slowly varying functions so that the integral can be approximated by $\sum_{j=1}^{k} \Phi(z_j)w^*(z_j)\Delta_j$. ince we have restricted our problem to stationary rkoff functions we will use the notation

$$r(x_0 \mid x, t; \lambda) \equiv r(x_0, t_0 \mid x, t_0 + t; \lambda) \qquad (17)$$

$$p(x_0 \mid x, t) \equiv p(x_0, t_0 \mid x, t_0 + t). \qquad (18)$$

he estimate $w^*(z)$ defined above can be written in form

$$w^*(z) = u/t$$

ere u is defined by (1), with $t_0 = 0$ and Φ a function x only, which is defined by

$$\Phi(x) = \begin{cases} \Delta^{-1} & \text{if} \quad |x - z| < \Delta/2, \\ 0 & \text{otherwise.} \end{cases} \qquad (19)$$

q. (6) becomes

$$\,\mid x, t; \lambda) = p(x_0 \mid x, t) - \frac{\lambda}{\Delta}\int_0^t dt' \int_{z-\Delta/2}^{z+\Delta/2} \cdot p(x_0 \mid x', t')r(x' \mid x, t - t'; \lambda)\,dx'. \qquad (20)$$

n the limit $\Delta \to 0$ this equation simplifies to the egral equation

$$\,\mid x, t; \lambda) = p(x_0 \mid x, t) - \lambda \int_0^t dt' \cdot p(x_0 \mid z, t')r(z \mid x, t - t'; \lambda) \qquad (21)$$

ich can be solved by taking Laplace transforms. f r_L and p_L denote the Laplace transforms of r and p, ,

$$r_L(x_0 \mid x, s; \lambda) = \int_0^\infty e^{-st} r(x_0 \mid x, t; \lambda)\,dt, \qquad (22)$$

obtains

$$x_0 \mid x, s; \lambda) = p_L(x_0 \mid x, s) - \lambda p_L(x_0 \mid z, s)r_L(z \mid x, s; \lambda). \qquad (23)$$

solve (23) for $r(x_0 \mid x, s; \lambda)$ we first put $x_0 = z$ and ain

$$r_L(z \mid x, s; \lambda) = p_L(z \mid x, s) - \lambda p_L(z \mid z, s)r_L(z \mid x, s; \lambda) \qquad (24)$$

and

$$r_L(z \mid x, s; \lambda) = p_L(z \mid x, s)/[1 + \lambda p_L(z \mid z, s)]. \qquad (25)$$

Substituting this result into (23) yields

$$r_L(x_0 \mid x, s; \lambda) = p_L(x_0 \mid x, s) - \lambda p_L(x_0 \mid z, s)p_L(z \mid x, s)/[1 + \lambda p_L(z \mid z, s)]. \qquad (26)$$

We denote by $\rho(x_0 \mid x, u, t)$ the joint probability density for x and u at t with fixed initial value of $x(\tau)$ $(x(0) = x_0)$, so that

$$r(x_0 \mid x, t; \lambda) = \int_0^\infty e^{-u\lambda}\rho(x_0 \mid x, u, t)\,du$$

and we denote by $\rho_L(x_0 \mid x, u, s)$ the Laplace transform of ρ with respect to t. We compute first $\rho_L(x_0 \mid x, u, s)$ by Laplace inversion of (26). To do this we have to split off the term which leads to a delta function in $\rho_L(x_0 \mid x, u, s)$ (unless its coefficient vanishes). This term corresponds to a delta function in $\rho(x_0 \mid x, u, t)$ and the coefficient of the delta function is the probability that $w^*(z) = 0$, which will occur with nonvanishing probability, for instance, if x_0 and x are both smaller or larger than z.

We, therefore, split off those terms in (26) which do not vanish in the limit $\lambda \to \infty$ and write

$$\begin{aligned} r_L(x_0 \mid x, s, \lambda) &= p_L(x_0 \mid x, s) \\ &\quad - p_L(x_0 \mid z, s)p(z \mid x, s)/p_L(z \mid z, s) \\ &\quad - \left[\frac{\lambda p_L(x_0 \mid z, s)p_L(z \mid x, s)}{1 + \lambda p_L(z \mid z, s)} - \frac{p_L(x_0 \mid z, s)p_L(z \mid x, s)}{p_L(z \mid z, s)}\right] \\ &= \frac{p_L(x_0 \mid x, s)p_L(z \mid z, s) - p_L(x_0 \mid z, s)p_L(z \mid x, s)}{p_L(z \mid z, s)} \\ &\quad + \frac{p_L(x_0 \mid z, s)p_L(z \mid x, s)}{p_L(z \mid z, s)(1 + \lambda p_L(z \mid z, s))}. \end{aligned} \qquad (27)$$

Taking the Laplace inverse with respect to λ we obtain[9]

$$\begin{aligned} \rho_L(x_0 \mid x, u, s) &= \frac{p_L(x_0 \mid x, s)p_L(z \mid z, s) - p_L(x_0 \mid z, s)p_L(z \mid x, s)}{p_L(z \mid z, s)}\,\delta_+(u) \\ &\quad + \frac{p_L(x_0 \mid z, s)p_L(z \mid x, s)}{p_L^2(z \mid z, s)}\,e^{-u/p_L(z \mid z, s)} \end{aligned} \qquad (28)$$

for $u \geq 0$ and zero otherwise, where $\delta_+(u)$ is defined by $\delta_+(u) = 0$ for $u \neq 0$ and

$$\int_0^\epsilon \delta_+(u)\,du = 1 \quad \text{for any} \quad \epsilon > 0. \qquad (29)$$

[9] Eq. (28) can be checked by comparing the moments of u with those obtained directly. See Appendix I.

The conditional probability of finding $w^*(z)t$ in the interval $(u, u + du)$, if $x(0) = x_0$ and $x(t) = x$ is known, is thus given by $\rho(x_0 \mid x, u, t)du/p(x_0 \mid x, t)$ where $\rho(x_0 \mid x, u, t)$ is to be obtained by Laplace inversion from $\rho_L(x_0 \mid x, u, s)$, given by (28). The (unconditional) probability of finding $w^*(z)t$ in the interval $(u, u + du)$ is given by

$$\rho(u, t) \equiv \iint_{-\infty}^{\infty} w(x_0)\rho(x_0 \mid x, u, t)\, dx_0\, dx$$

and is to be obtained by Laplace inversion of

$$\begin{aligned}\rho_L(u, s) &\equiv \iint_{-\infty}^{\infty} w(x_0)\rho_L(x_0 \mid x, u, s)\, dx_0\, dx \\ &= \left(\frac{1}{s} - \frac{w(z)}{s^2 p_L(z \mid z, s)}\right) \delta_+(u) \\ &\quad + \frac{w(z)}{s^2 p_L^2(z \mid z, s)}\, e^{-u/p_L(z \mid z, s)}. \qquad (30)\end{aligned}$$

This inversion will in general be too complicated to perform exactly, but an asymptotic evaluation for large t can be carried through. We expect that the quantity $(u - wt)/\sqrt{t}$ becomes normally distributed in this limit. We thus consider the distribution

$$\begin{aligned}F(v, t) &\equiv \text{prob } \{(u - wt)/\sqrt{t} \leq v\} \\ &= \begin{cases} \displaystyle\int_0^{v\sqrt{t} + wt} du \left\{\frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} \rho_L(u, s)e^{st}\, ds\right\} & \text{for } v \geq -w\sqrt{t} \\ 0 \text{ for } v < -w\sqrt{t} \end{cases} \qquad (31)\end{aligned}$$

where the path of integration has to be taken to the right of singularities of ρ_L. Using (30) and interchanging the order of integration yields

$$\begin{aligned}F(v, t) &= \frac{1}{2\pi i}\int_{-i\infty+\gamma}^{+i\infty+\gamma} \left\{\frac{1}{s} - \frac{w}{s^2 p_L}\, e^{-(v\sqrt{t} + wt)/p_L}\right\} e^{st}\, ds \qquad (32) \\ &= 1 - \frac{w}{2\pi i}\int_{-i\infty+\gamma}^{+i\infty+\gamma} \frac{ds}{s^2 p_L} \\ &\quad \cdot \exp\, [(-vs\sqrt{t} + s^2 t(p_L - w/s))/(w + s(p_L - w/s)]\end{aligned}$$

for $v \geq - w \sqrt{t}$, where w and p_L stand for $w(z)$ and $p_L(z \mid z, s)$, respectively. With $\zeta = - is \sqrt{t}$ and $\tau(z, s) \equiv p_L - w/s$ we write this result in the form

$$\begin{aligned}F(v, t) &= 1 - \frac{w}{2\pi i}\int_{-\infty}^{\infty} \frac{d\zeta}{\zeta}\, \frac{1}{w + i\zeta\tau(z, i\zeta/\sqrt{t})/\sqrt{t}}. \qquad (33) \\ &\cdot \exp\{[-i\zeta v - \zeta^2\tau(z, i\zeta/\sqrt{t})]/[w + i\zeta\tau(z, i\zeta/\sqrt{t})/\sqrt{t}]\}\end{aligned}$$

where the path of integration has to be taken below any singularities of the integrand. If

$$\lim_{s \to 0} \tau(z, s) = \tau(z) \qquad (34)$$

exists at least for Re $s \geq 0$ and is finite and larger than zero and if the limit $t \to \infty$ and the integration can be interchanged, then $F(v, t)$ approaches the normal distribution for $v \geq - w \sqrt{t}$, *i.e.*,

$$F(v, \infty) = [4\pi\tau(z) \cdot w(z)]^{-1/2} \int_{-\infty}^{v} e^{-\eta^2/4w(z)\tau(z)}\, d\eta. \qquad (35)$$

For the existence of the limit $\tau(z)$ it is sufficient that the stationary distribution is approached sufficiently fast so that

$$\int_0^{\infty} \mid p(z \mid z, t) - w(z) \mid dt \quad \text{exists.} \qquad (36)$$

The significance of the condition $\tau(z) > 0$ can be seen in various ways. We note first that the unconditional first and second moments of u are, according to (51) in Appendix II,

$$\bar{u} = wt \qquad (37)$$

$$\begin{aligned}\langle u^2 \rangle_{\text{av}} &= 2w \int_0^t dt_2 \int_0^{t_2} p(z \mid z, t_2 - t_1)\, dt_1 \\ &= 2w \int_0^t dt_2 \int_0^{t_2} p(z \mid z, t')\, dt' \\ &= 2w\left\{t\int_0^t p(z \mid z, t')\, dt' - \int_0^t t_2 p(z \mid z, t_2)\, dt_2\right\}. \qquad (38)\end{aligned}$$

One has thus

$$\langle u^2 - \bar{u}^2 \rangle_{\text{av}} = 2w \int_0^t \{p(z \mid z, t') - w\}(t - t')\, dt' \qquad (39)$$

or

$$t^{-1}\langle u^2 - \bar{u}^2 \rangle_{\text{av}} = 2w\int_0^t \{p(z \mid z, t') - w\}\left(1 - \frac{t'}{t}\right) dt'. \quad (40)$$

In the limit $t \to \infty$, the second factor in the integrand is merely a convergence creating factor, so that if the integral converges with the first factor alone, *i.e.*, *a fortiori* if the condition (36) is fulfilled

$$\begin{aligned}\lim_{t \to \infty} t^{-1}\langle u^2 - \bar{u}^2 \rangle_{\text{av}} &= 2w(z) \int_0^{\infty} \{p(z \mid z, t') - w(z)\}\, dt' \\ &= 2w(z)\tau(z). \qquad (41)\end{aligned}$$

This shows that for $w(z) \neq 0$, $\tau(z)$ must be at least nonnegative.

From (26) or (30) one sees that

$$\int_0^{\infty} e^{-st}\, \text{prob } \{u \neq 0\}\, dt = w/s^2 p_L. \qquad (42)$$

We can consider prob $\{u \neq 0\}$ also as the probability that the time ϑ at which the first contribution to u occurs[10] is smaller than t. The first moment $\bar{\vartheta}$ is, therefore,

[10] For a continuous process, this is the first-passage time.

$$\bar{\vartheta} = -\left[\frac{d}{ds}\left(\frac{w}{sp_L}\right)\right]_{s=0}$$

$$= w\left[\frac{p_L + sdp_L/ds}{s^2 p_L^2}\right]_{s=0}$$

$$= w^{-1}[p_L - w/s]_{s=0}$$

$$= w^{-1}\tau(z). \qquad (43)$$

iis shows that $\tau(z)$ is finite and positive as required, if $\neq 0$ and if the average time at which the first conoution to u occurs is finite and different from zero.

Appendix I

ľo derive (5) more formally we use the trivial identity

$$\text{p}\left\{-\lambda\int_{t_0}^{t}\Phi(X(\tau),\tau)\,d\tau\right\} = 1 - \lambda\int_{t_0}^{t} dt'\Phi(X(t'),t')$$
$$\cdot\exp\left\{-\lambda\int_{t_0}^{t}\Phi(X(\tau),\tau)\,d\tau\right\}. \qquad (44)$$

averaging both sides with initial and end point of $t)$ fixed, and multiplying by $p(X_0, t_0 \mid X, t)$ one obtains ɔm the definition (4)

$$X_0, t_0 \mid X, t; \lambda)$$
$$= p(X_0, t_0 \mid X, t)\Bigg(1 - \lambda\int_{t_0}^{t} dt'$$
$$\cdot\left[\Phi(X(t'),t')\exp\left\{-\lambda\int_{t_0}^{t'}\Phi(X(\tau),\tau)\,d\tau\right\}\right|$$
$$\left.\cdot X(t_0) = X_0, X(t) = X\right]_{\text{av}}\Bigg). \qquad (45)$$

the second term, we now introduce a third condition $(t') = X'$ and compensate for this by multiplicatiom with $X_0, t_0; X, t \mid X', t')dX'$ and integration over X', where $X_0, t_0; X, t \mid X', t')dX'$ is defined as the probability at $X(t')$ in dX' at X' for a path with fixed initial and d point $X(t_0) = X_0$ and $X(t) = X$, respectively. The cond term thus becomes

$$dt'\int dX'\,p(X_0, t_0; Xt \mid X', t')$$
$$\left[\Phi(X(t'),t')\exp\left\{-\lambda\int_{t_0}^{t'}\Phi(X(\tau),\tau)\,d\tau\right\}\right|$$
$$\left.\cdot X(t_0) = X_0, X(t') = X', X(t) = X\right]_{\text{av}}. \qquad (46)$$

e can now take $\Phi(X(t'), t')$ out of the average symbol $\Phi(X', t')$, since $X(t') = X'$ is held fixed. We also can nit the condition $X(t) = X$ in the average symbol nce, by virtue of the Markoff property, $X(t)$ is statistilly independent of values of $X(t'')$ for $t'' < t'$, if $X(t')$ is ld fixed. Also from the Markoff property and the defition of conditional probabilities follows for $t_0 \leq t' \leq t$,

$$X_0, t_0; Xt \mid X', t') = p(X_0, t_0 \mid X', t')$$
$$\cdot p(X', t' \mid X, t)/p(X_0, t_0 \mid X, t). \qquad (47)$$

We thus have

$$r(X_0, t_0 \mid X, t; \lambda) = p(X_0, t_0 \mid X, t) - \lambda\int_{t_0}^{t'} dt'\int dX'$$
$$\left[\exp\left\{-\lambda\int_{t_0}^{t'}\Phi(X(\tau),\tau)\,d\tau\right\}\right|$$
$$\left.\cdot X(t_0) = X_0, X(t') = X'\right]_{\text{av}} p(X_0, t_0 \mid X', t')$$
$$\cdot\Phi(X', t')p(X', t' \mid X, t) \qquad (48)$$

and using the definition of $r(X_0, t_0 \mid X', t', \lambda)$ on the right-hand side, we obtain (5). By a similar argument starting with the identity

$$\exp\left\{-\lambda\int_{t_0}^{t}\Phi(X(\tau),\tau)\,d\tau\right\} = 1 - \lambda\int_{t_0}^{t} dt'\,\Phi(X(t'),t')$$
$$\cdot\exp\left\{-\lambda\int_{t'}^{t}(X(\tau),\tau)\,d\tau\right\} \qquad (49)$$

one obtains (6).

Appendix II

Eq. (28) can be checked directly by the method of moments. Since for integer $n \geq 1$

$$u^n = \int\overset{t}{\underset{0}{\cdots}}\int\prod_{j=1}^{n}\left\{dt_j\,\frac{1}{\Delta}\int_{z-\Delta/2}^{z+\Delta/2}\delta(x(t_j) - z_j)\,dz_j\right\}$$
$$= n!\int_0^t dt_n\int_0^{t_n} dt_{n-1}\cdots\int_0^{t_2} dt_1\,\Delta^{-n}$$
$$\cdot\prod_{j=1}^{n}\int_{z-\Delta/2}^{z+\Delta/2}\delta(x(t_j) - z_j)\,dz_j \qquad (50)$$

one has

$$\langle u^n \mid x(t_0) = x_0, x(t) = x\rangle_{\text{av}}$$
$$= n!\int_0^t dt_n\int_0^{t_n} dt_{n-1}\cdots\int_0^{t_2} dt_1\,\Delta^{-n}\iint_{z-\Delta/2}^{z+\Delta/2} p(x_\circ \mid z_1, t_1)$$
$$\cdot\prod_{j=1}^{n-1} p(z_j \mid z_{j+1}, t_{j+1} - t_j)p(z_n \mid x, t - t_n)\prod_{1}^{n} dz_j$$
$$= n!\int_0^t dt_n p(z \mid x, t - t_n)\int_0^{t_n} dt_{n-1}$$
$$\cdot p(z \mid z, t_n - t_{n-1})\cdots\int_0^{t_2} p(z \mid z, t_2 - t_1)$$
$$\cdot p(x_0 \mid z, t_1). \qquad (51)$$

Taking Laplace transforms one has

$$\int_0^\infty e^{-st}\,dt\,\langle u^n \mid x(t_0) = x_0, x(t) = x\rangle_{\text{av}}$$
$$= n!\,p_L(x_0 \mid z, s)p_L(z \mid x, s)p_L(z \mid z, s)^{n-1}. \qquad (52)$$

This is in agreement with Laplace transforms of the moments obtained by integration with respect to u from (28).

A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena —Part II, Examples*

ARNOLD J. F. SIEGERT†

Summary—The method of Part I is applied to the problem of finding the probability distribution of $u \equiv \int_0^t K(\tau)x^2(\tau)\,d\tau$, where $K(\tau)$ is a given function and $x(\tau)$ is the Uhlenbeck process. The earlier methods of Kac and the author yielded the characteristic function of this distribution as the reciprocal square root of the Fredholm determinant D of an integral equation. The present method yields a second-order linear differential equation with initial condition only for D as function of t. For the special cases $K(\tau) = 1$ and $K(\tau) = e^{-\alpha\tau}$ the characteristic function is obtained in closed form.

In Section III, we have verified directly from the integral equation the differential equation for D and some relations between D and the initial and end point values of the Volterra reciprocal kernel which appear in the joint characteristic function for u, $x(0)$ and $x(t)$.

Section I

In previous papers,[1] Darling and the author derived two integral equations for a function closely related to the characteristic function of the probability distribution of the functional $\int_{t_0}^t \phi(X(\tau), \tau)\,d\tau$, where the components $x_j(\tau)$ of $X(\tau)$ form a Markoff process and $\phi(X, \tau)$ is a given function of X and τ. For an important class of Markoff processes, the solution of these integral equations was shown to be the principal solution of a partial differential equation similar to the Fokker-Planck equation. We also derived an integral equation relating two solutions of the problem with two functions, $\phi_2(X, \tau)$ and $\phi_1(X, \tau)$, which can be used for a perturbation calculation to obtain solutions of the problem when the solution for $\phi_1(X, \tau)$ is known, and the solution for a function $\phi_2(X, \tau)$ "in the neighborhood" of $\phi_1(X, \tau)$ is desired.

Since the integral equations as well as the differential equations derived in this previous paper are rather formidable, it seemed of interest to consider first some cases for which the problem could be solved or at least be reduced to an integral equation in a single variable by an older method.[2,3] In the case $\phi(X, \tau) = K(\tau)x^2(\tau)$, where $x(\tau)$ is a Gaussian random function with arbitrary autocorrelation function $\rho(\tau)$, the older method applies and the problem can be reduced to the problem of solving either the homogeneous integral equation[2]

$$\int_{t_0}^{t} \rho(\tau_1 - \tau)K(\tau)\varphi(\tau)\,d\tau = \lambda\varphi(\tau_1) \tag{1}$$

or the inhomogeneous integral equation[3]

$$g_\lambda(\tau_1, \tau_2) + 2\lambda \int_{t_0}^{t} \rho(\tau_1 - \tau) \cdot K(\tau)g_\lambda(\tau, \tau_2)\,d\tau = \rho(\tau_1 - \tau_2). \tag{2}$$

If, specially, the function $x(\tau)$ is a component $x_1(\tau)$ of a Markoffian Gaussian process, the method referred to[1] applies and leads to the partial differential equation [(15) of Part I]:

$$(L - \lambda K(t)x_1^2)r = \frac{\partial r}{\partial t} \tag{3}$$

if the transition probability density $p(X_0 \mid X, t)$ is the principal solution of the differential equation

$$Lp = \frac{\partial p}{\partial t} \tag{4}$$

where L operates on the components of X. One sees easily that in the case of Gaussian $p(X_0 \mid X, t)$ the additional term in (3) does not seriously complicate the differential equation and that r is still of the Gaussian form. For the coefficients of this Gaussian, one obtains a system of differential equations of first order and second degree with t as the independent variable with initial conditions only. The equivalence of this system in which t appears as the independent variable and the integral equation (2) in which t appears only parametrically is not trivial even though (2) can be reduced to a differential equation with appropriate boundary conditions when $\rho(\tau)$ is the auto-correlation function of a Markoffian Gaussian process. We have, therefore, in Section II worked out in detail the case of the one-dimensional, Markoffian Gaussian random process (Uhlenbeck process).[4] As special examples, we give in closed form the characteristic functions of $\int_0^t x^2(\tau)d\tau$ and $\int_0^t e^{-\alpha\tau}x^2(\tau)d\tau$. The latter represents the output of a receiver consisting of single-tuned IF and audio stage with quadratic detector, with

* Manuscript received by the PGIT, August 13, 1956. Most of this work was done while the author was a consultant for The RAND Corp., Santa Monica, Calif.

† Dept. of Physics, Northwestern University, Evanston, Ill.

[1] D. A. Darling and A. J. F. Siegert, "On the Distribution of Certain Functionals of Markoff Processes," The RAND Corp., Paper P-429; October 31, 1953, and Part I of this paper (P-738), "A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena;" September, 1955. Darling and Siegert, "On the distribution of certain functionals of Markoff chains and processes," *Proc. Natl. Acad. Sci.*, vol. 42, pp. 525–529; August, 1956.

[2] M. Kac and A. J. F. Siegert, "Note on the theory of noise in receivers with square law detector," *Phys. Rev.*, vol. 70, p. 449; September, 1946.

Kac and Siegert, "On the theory of noise in radio receivers with square law detector," *J. Appl. Phys.*, vol. 18, pp. 383–397; April, 1947.

[3] A. J. F. Siegert, "Passage of stationary processes through linear and non-linear devices," IRE Trans., vol. PGIT-3, pp. 4–25; March, 1954. (The RAND Corp., (P-419); October 29, 1953.)

[4] Analogous problems for the Wiener process have been treated independently by R. Deutsch, "Piecewise quadratic detector," 1956 IRE Convention Record, Part 4, pp. 15–20.

te noise input turned on a time t before observation. Section III, we have shown for this case how the system differential equations originally obtained by the thod[1] follows directly from (2). The purpose of this ivation was primarily to show that some of the equans derived by the method[1] remain valid when the o-correlation function of the Uhlenbeck process $_1 - \tau) = \exp(-\beta \mid \tau_1 - \tau \mid)]$ is replaced by a general ction $h(\tau_1, \tau)$. We found, *e.g.*, a simpler expression for the racteristic function $f \equiv \langle \exp[-\lambda \int_0^t K(\tau)x^2(\tau)d\tau]\rangle_{\text{av}}$ ere $x(\tau)$ is a Gaussian random function with arbitrary rrelation function $\rho(\tau_1 - \tau)$. We had shown[3] that the ction f can be expressed in terms of the trace of the ution $g_\lambda(\tau_1, \tau_2)$ of the integral equation (2) (which pends parametrically on t) by

$$f = \exp\left[-\int_0^\lambda d\kappa \int_0^t K(\tau)g_\kappa(\tau, \tau)\, d\tau\right].$$

is equation is suggested by the well-known expression the Fredholm determinant in terms of the Volterra iprocal function. We now obtained the simpler expssion

$$f = \exp\left[-\lambda \int_0^t K(t')g_\lambda(t', t')\, dt'\right]$$

ere $g_\lambda(\tau_1, \tau_2)$ is the solution of (2) with the upper limit placed by t' so that $g_\lambda(t', t')$ depends on t' implicitly o. [See (58).]

The results[1] and the present paper raise an interesting estion for further investigation. In Kac and Siegert,[2] e fact that the problem of the probability distribution a quadratic integral form of a Gaussian process $x(t)$ uld be reduced to the solution of an integral equation volving as variable only the time was clearly a conseence of the fact that the joint probability distribution $x(t_1), x(t_2) \cdots x(t_n)$ is the exponential function of a adratic form. The results[1] show that this simplification n also be understood as a consequence of the fact that e differential equation (4) is not essentially complicated the addition of a quadratic term. One may thus expect find for Markoff processes other than the Gaussian ocesses certain functionals for which the problem of ding the characteristic function reduces to differential uations with only the time as independent variable.

Section II

In this section we will apply the method[1] to the problem evaluating the function

$$\begin{aligned} &x_0 \mid x, t, \lambda) \\ &= \left\langle \exp\left[-\lambda \int_0^t K(\tau)x^2(\tau)\, d\tau\right] \middle| x(0) = x_0, x(t) = x \right\rangle_{\text{av}} \\ &\qquad \cdot p(x_0 \mid x, t) \qquad (5) \end{aligned}$$

here $x(\tau)$ is the stationary one-dimensional Markoffian aussian process (Uhlenbeck process), which is described the transition probability density

$$p(x_0 \mid x, t) = [2\pi(1 - e^{-2\beta t})]^{-1/2} \exp\left\{-\frac{(x - x_0e^{-\beta t})^2}{2(1 - e^{-2\beta t})}\right\} \qquad (6)$$

with constant β, $K(\tau)$ is a given function, λ is to be chosen positive if $K(\tau) \geq 0$ and negative imaginary if $K(\tau)$ can assume negative values. The symbol $\langle | \rangle_{\text{av}}$ denotes the average of the functional to the left of the vertical line under the conditions written to the right of the vertical line. The extension of the result to, *e.g.*, the characteristic function for $\int_0^t K(\tau) \sum_j x_j^2(\tau)\, d\tau$ where $x_j(\tau)$ are independent Uhlenbeck processes is trivial.

This problem can be treated by the method[2] and the result can be brought into a slightly more convenient form by an extension of this method.[3]

It is convenient to work with the function

$$\begin{aligned} \hat{r}(\eta \mid \zeta, t, \lambda) &\equiv (2\pi)^{-1/2} \iint_{-\infty}^{\infty} \exp\left(-\frac{x_0^2}{2} + i\eta x_0 + i\zeta x\right) \\ &\qquad \cdot r(x_0 \mid x, t, \lambda)\, dx\, dx_0 \qquad (7) \\ &= \left\langle \exp\left\{i[\eta x(0) + i\zeta x(t)] - \lambda \int_0^t K(\tau)x^2(\tau)\, d\tau\right\}\right\rangle_{\text{av}} \end{aligned}$$

from which $r(x_0 \mid x, t, \lambda)$ is easily obtained by Fourier inversion and multiplication by $\sqrt{2\pi}\ e^{x_0^2/2}$. Note that the characteristic function for the distribution of $\int_0^t K(\tau)x^2(\tau)d\tau$ without initial and end conditions is simply obtained by choosing $\eta = \zeta = 0$.

The result obtained by using the methods[2,3] is

$$\begin{aligned} \hat{r}(\eta \mid \zeta, t, \lambda) = \exp\Bigg\{&-\int_0^\lambda d\kappa \int_0^t K(\tau)g_\kappa(\tau, \tau)\, d\tau \\ &-\frac{1}{2}\left(\eta^2 g_\lambda(0, 0) + 2\eta\zeta g_\lambda(0, t) + \zeta^2 g_\lambda(t, t)\right)\Bigg\} \qquad (8) \end{aligned}$$

where $g_\lambda(\tau_1, \tau_2)$ is the solution of the integral equation

$$\begin{aligned} g_\lambda(\tau_1, \tau_2) + 2\lambda \int_0^t K(\tau)\rho(\tau_1 - \tau) \\ \cdot g_\lambda(\tau, \tau_2)\, d\tau = \rho(\tau_1 - \tau_2). \qquad (9) \end{aligned}$$

(See Appendix II.)

This result is valid for any stationary Gaussian process with auto-correlation function $\rho(\tau)$. The variable τ_2 appears only as a parameter in the integral equation. If $x(\tau)$ represents the output of a network with lumped circuit and with white noise input, the integral equation reduces to a differential equation. For the Uhlenbeck process, one has specially $\rho(\tau) = e^{-\beta|\tau|}$, and one would reduce the integral equation to a second-order linear differential equation with appropriate boundary conditions with τ_1 as the independent variable.

It is interesting to see how the new method leads to a Riccati equation for $g_\lambda(t, t)$, with t as the independent variable. This Riccati equation is of course also equivalent to a second-order linear differential equation, which, however, is not the same as the differential equation obtained for $g_\lambda(\tau_1, \tau_2)$ as function of τ_1.

The new method leads to the differential equation

$$\frac{\partial r}{\partial t} = \beta\left\{\frac{\partial^2 r}{\partial x^2} + \frac{\partial(xr)}{\partial x}\right\} - \lambda K x^2 r \tag{10}$$

Since p satisfies

$$\frac{\partial p}{\partial t} = \beta\left\{\frac{\partial^2 p}{\partial x^2} + \frac{\partial(xp)}{\partial x}\right\}. \tag{11}$$

From this one obtains

$$\frac{\partial \hat{r}}{\partial t} = -\beta\left(\zeta \frac{\partial \hat{r}}{\partial \zeta} + \zeta^2 \hat{r}\right) + \lambda K \frac{\partial^2 \hat{r}}{\partial \zeta^2}. \tag{12}$$

Making the Ansatz

$$\hat{r} = f \exp\left[-\tfrac{1}{2}(\sigma_0 \eta^2 + 2\sigma\zeta\eta + \sigma_1\zeta^2)\right] \tag{13a}$$

one obtains (with dots indicating differentiation with respect to t)

$$\frac{\dot{f}}{f} - \tfrac{1}{2}(\dot{\sigma}_0\eta^2 + 2\dot{\sigma}\zeta\eta + \dot{\sigma}_1\zeta^2) = -\beta[\zeta(-\sigma\eta - \sigma_1\zeta) + \zeta^2] + \lambda K[(\sigma\eta + \sigma_1\zeta)^2 - \sigma_1] \tag{13b}$$

and by comparing coefficients

$$\frac{d \ln f}{dt} = -\lambda K \sigma_1 \tag{14a}$$

$$\frac{d\sigma_1}{dt} = 2\beta(1 - \sigma_1) - 2\lambda K \sigma_1^2 \tag{14b}$$

$$\frac{d\sigma_0}{dt} = -2\lambda K \sigma^2 \tag{14c}$$

$$\frac{d \ln \sigma}{dt} = -\beta - 2\lambda K \sigma_1. \tag{14d}$$

Since $r(x_0 \mid x, 0, \lambda) = \delta(x - x_0)$, we have

$$\hat{r}(\eta \mid \zeta, 0, \lambda) = (2\pi)^{-1/2} \iint_{-\infty}^{\infty} \exp\left(-\frac{x_0^2}{2} + i\eta x_0 + i\zeta x\right) \delta(x - x_0)\, dx\, dx_0 \tag{15}$$

$$= (2\pi)^{-1/2} \int_{-\infty}^{\infty} \exp\left(-\frac{x_0^2}{2} + i(\eta + \zeta)x_0\right) dx_0$$

$$= \exp\left(-\tfrac{1}{2}(\eta + \zeta)^2\right).$$

The initial conditions for the differential equations (14a) to (14d) are, therefore,

$$f(0) = \sigma_0(0) = \sigma_1(0) = 1. \tag{16}$$

We note that σ_0, σ, and f are obtained by quadratures, if σ_1 has been found, which means that $g_\lambda(0, 0)$, $g_\lambda(0, t)$ and $\int_0^\lambda d\kappa \int_0^t K(\tau) g_\kappa(\tau, \tau) d\tau$ are obtained by quadratures from $g_\lambda(t, t)$. It should be remembered that these relations need hold only for the special choice $\rho(\tau) = e^{-\beta|\tau|}$ since a stationary Gaussian process is Markoffian only if it has this special auto-correlation function, and (10) is based on the Markoff property. We will show in the following Section III, however, that (14a) and (14c) are independent of this special form of ρ.

It will generally be more convenient to convert the Riccati equation (10b) into a second-order linear differential equation. With the substitutions

$$dx = 2\lambda K\, dt \tag{17}$$

and

$$\sigma_1 = u'/u \tag{18}$$

where prime denotes differentiation with respect to x we get

$$\frac{d\sigma_1}{2\lambda K\, dt} = \frac{\beta}{\lambda K}(1 - \sigma_1) - \sigma_1^2 \tag{19}$$

$$\left(\frac{u'}{u}\right)' = \frac{\beta}{\lambda K}\left(1 - \frac{u'}{u}\right) - \left(\frac{u'}{u}\right)^2 \tag{20}$$

or

$$\lambda\beta^{-1} K(t(x))\, u'' + u' - u = 0. \tag{21}$$

One initial condition is obtained from $\sigma_1(0) = 1$

$$u'(x_0) = u(x_0) \tag{22a}$$

which simplifies to

$$u''(x_0) = 0 \quad \text{if} \quad K(t(x_0)) \neq 0 \tag{22b}$$

where x_0 is the value assumed by x for $t = 0$. This determines $u(x)$ except for a constant factor which is irrelevant since all results depend only on $\sigma_1 = u'/u$. Of special interest is the characteristic function for the unconditional distribution, f. We obtain from (14a)

$$\frac{d \ln f}{2\lambda K\, dt} = \frac{d \ln f}{dx} = -\frac{1}{2}\sigma_1 = -\frac{1}{2}\frac{d \ln u}{dx} \tag{23}$$

or

$$f = \sqrt{\frac{u(x_0)}{u(x)}}. \tag{24}$$

Since $u(x_0)$ is essentially the Wronskian, it can always be written in a convenient form (see Appendix 1).

The present method thus presents the characteristic function in terms of the solution of a differential equation with initial condition rather than through an eigenvalue problem or an inhomogeneous integral equation.

For σ and σ_0 we obtain from (14c), (14d), and (16)

$$d \ln \sigma = -\beta\, dt - \sigma_1\, dx \tag{25}$$

$$\ln \sigma = -\beta t(x) - \ln u + \text{const} \tag{25'}$$

$$\sigma = e^{-\beta t}\, u(x_0)/u(x) \tag{26}$$

and

$$\frac{d\sigma_0}{dx} = -\sigma^2 = -e^{-2\beta t(x)}\, u^2(x_0)/u^2(x) \tag{27}$$

or

$$\sigma_0 = 1 - u^2(x_0) \int_{x_0}^{x} e^{-2\beta t(y)}\, u^{-2}(y)\, dy. \tag{28}$$

Kac and Siegert,[2] solutions in terms of infinite pro-ts were given for the two cases $K(\tau) = 1$, and -) $= e^{-\alpha\tau}$ with $t = \infty$. We will give here the two utions in closed form with the second case for general t.

Case 1: $K(t) = 1$.

$$\lambda\beta^{-1}u'' + u' - u = 0 \tag{29}$$

h

$$x = 2\lambda t. \tag{30}$$

e general solution is

$$u = e^{-\beta x/2\lambda}\,(ae^{kx} + be^{-kx}) \tag{31}$$

h

$$\kappa = \sqrt{(\beta/2\lambda)^2 + \beta/\lambda} \tag{32}$$

e initial condition (22a) requires

$$\frac{a - b}{a + b} = \kappa^{-1}(1 + \beta/2\lambda). \tag{33}$$

e thus have from (24)

$$= e^{\beta x/4\lambda}\left(\frac{a + b}{ae^{\kappa x} + be^{-\kappa x}}\right)^{1/2}$$

$$= e^{\beta x/4\lambda}\left(\frac{a + b}{(a + b)\cosh \kappa x + (a - b)\sinh \kappa x}\right)^{1/2}$$

$$= e^{\beta x/4\lambda}\,(\cosh \kappa x + \kappa^{-1}(1 + \beta/2\lambda)\sinh \kappa x)^{-1/2}. \tag{34}$$

ith

$$\eta \equiv \sqrt{1 + 4\lambda/\beta} \tag{35}$$

is becomes[5]

$$f = e^{\beta t/2}\,(\cosh \beta\eta t + \eta^{-1}\,(1 + 2\lambda/\beta)\sinh \beta\eta t)^{-1/2}. \tag{36}$$

he roots $\lambda^{(m)}$ of f are determined by the roots η_ν of

$$\mathrm{th}\beta\eta t = -\frac{\eta}{1 + 2\lambda/\beta} = -\frac{2\eta}{1 + \eta^2} \tag{37}$$

th

$$\lambda^{(m)} = \frac{\beta}{4}\,(\eta_m^2 - 1). \tag{38}$$

Kac and Siegert,[2] the meaning of α and β is interanged and λ_ν is in our present notation given by

$$\lambda_m = -\frac{1}{2\lambda^{(m)}t} \tag{39}$$

that we have

$$\lambda_\nu = \frac{2}{\beta t(1 - \eta_m^2)}. \tag{40}$$

ith $\eta = iy$ and $\eta_m = -iy_m$ we then have agreement ith equations (7.35) and (7.36) of Kac and Siegert.[2]

Case 2: $K(t) = e^{-\alpha t}$.

In this case we get from (17)

$$x = 2\lambda\alpha^{-1}e^{-\alpha t} \tag{41}$$

since it turns out to be convenient to choose the integration constant equal to zero. We thus have

$$x_0 = -2\lambda\alpha^{-1} \tag{42}$$

and

$$\lambda K(t(x)) = -1/2\ \alpha x. \tag{43}$$

Eq. (21) becomes

$$u'' - \frac{2\beta}{\alpha x}(u' - u) = 0 \tag{44}$$

and has the solution[6]

$$u = x^{p/2}Z_p(\sqrt{\gamma x}) \tag{45}$$

where $Z_p(y)$ is a Bessel function of order

$$p = 1 + (\gamma/4) \tag{46a}$$

and

$$\gamma = 8\beta/\alpha. \tag{46b}$$

With the aid of the identity[7]

$$\frac{d}{dx}\,[x^{p/2}Z_p(\sqrt{\gamma x})] = \tfrac{1}{2}\sqrt{\gamma}\,x^{(p-1)/2}Z_{p-1}(\sqrt{\gamma x}) \tag{47}$$

the initial condition is seen to be satisfied by

$$u(x) = x^{p/2}[N_{p-2}(\sqrt{\gamma x_0})J_p(\sqrt{\gamma x}) - J_{p-2}(\sqrt{\gamma x_0})N_p(\sqrt{\gamma x})] \tag{48}$$

where J_p is the ordinary Bessel function and N_p is the Neumann function

$$N_p = (J_p \cos \pi p - J_{-p})/\sin \pi p. \tag{49}$$

The function $u(x_0)$ must simplify since it is essentially a Wronskian and we have using (47)

$$x_0^{(p-2)/2}u(x_0) = \frac{2}{\sqrt{\gamma}}\frac{d}{dx_0}[x_0^{(p-1)/2}N_{p-1}x_0^{p/2}J_p - x_0^{(p-1)/2}J_{p-1}x_0^{p/2}N_p]$$

$$= \frac{2}{\sqrt{\gamma}}\frac{d}{dx_0}\left\{x_0^{p-(1/2)}\,\frac{2}{\pi\sqrt{\gamma x_0}}\right\} \tag{50}$$

where the second equation follows from an identity.[8] We thus get

$$u(x_0) = 4\gamma^{-1}\pi^{-1}(p - 1)x_0^{(p-2)/2} = \pi^{-1}x_0^{(p-2)/2} \tag{51}$$

and from (24) for the unconditional characteristic function

$$f(t) = \{\pi x_0^{(2-p)/2}x^{p/2}[N_{p-2}(\sqrt{\gamma x_0})J_p(\sqrt{\gamma x}) - J_{p-2}(\sqrt{\gamma x_0})N_p(\sqrt{\gamma x})]\}^{-1/2} \tag{52}$$

[5] In Kac and Siegert (footnote 2), f^2 was given in product form, nce we treated there the envelope detector.

[6] E. Jahnke and F. Emde, "Tables of Functions," Dover Publications, New York, N.Y., 4th ed., p. 146, sec. 7, second equation; 1945.

[7] *Ibid.*, p. 145, sec. 5, fourth equation.

[8] *Ibid.*, p. 144, sec. 4, third equation.

where x, x_0, p and γ are given by (41), (42), (46a), and (46b).

For $\alpha \to 0$ this must reduce to (36). For $t \to \infty$, $x \to 0$ and the first term in the bracket becomes negligible. We then have

$$x^{p/2}N_p(\sqrt{\gamma x}) \cong -\pi^{-1}(\tfrac{1}{2}\sqrt{\gamma})^{-p}\ \Gamma(p) \tag{53}$$

$$f(\infty) = \{\gamma^{-1}2^p(\gamma x_0)^{(2-p)/2}J_{p-2}(\sqrt{\gamma x_0})\ \Gamma(p)\}^{-1/2}. \tag{54}$$

The product form[2] is then checked by using the Weierstrass product[9]

$$(\gamma x_0)^{(2-p)/2}J_{p-2}(\sqrt{\gamma x_0}) = 2^{2-p}\prod_n\left(1 - \frac{\gamma x_0}{y_n^2}\right)\Big/\Gamma(p-1) \tag{55}$$

which yields

$$f(\infty) = \prod_n\left(1 - \frac{\gamma x_0}{y_n^2}\right)^{-1/2} = \prod_n\left(1 + \frac{4\beta\lambda}{\alpha^2 y_n^2}\right)^{-1/2} \tag{56}$$

where the numbers y_n are the roots of $J_{p-2}(y)$.

In Kac and Siegert,[2] the probability density $P(V)$ was obtained for the random variable $\beta\int_0^\infty e^{-\beta\tau}\ (x_1^2(\tau) + x_2^2(\tau))d\tau$, where $x_1(\tau)$ and $x_2(\tau)$ are two independent Uhlenbeck processes with $\overline{x_1^2} = \overline{x_2^2} = 1/2$ and autocorrelation function $e^{-\alpha|\tau|}$. We get for this the result (after interchanging α and β to conform with the notation[2]):

$$\begin{aligned} P(V) &= \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} e^{\lambda V/2\beta}f^2(\infty)\ d\lambda/2\beta \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\zeta V}\ d\zeta\prod_n\left(1 - \frac{8i\alpha\zeta}{\beta y_n^2}\right)^{-1} \end{aligned} \tag{57}$$

where yn is the nth root of $J_{(2\alpha|\beta)-1}(y) = 0$ in agreement with equations (4.45) and (7.19) of Kac and Siegert.[2]

Section III

It seemed to interest to verify the differential equations (14a) to (14d) directly from the integral equation (9) and to show that (14a) and (14c) remain valid when ρ is replaced by a general symmetric kernel $h(\tau_1, \tau_2)$. We will here only outline the derivation; the details are given in a RAND Corp. paper.[10]

From the integral equation

$$g(\tau_1, \tau_2) + 2\lambda\int_0^t h(\tau_1, \tau)K(\tau)g(\tau, \tau_2)\ d\tau = h(\tau_1, \tau_2) \tag{58}$$

one obtains an integral equation for $\partial g(\tau_1, \tau_2)/\partial t$ and an integral equation for $g(\tau_1, t)$. Comparing these two integral equations, whose solutions can be shown to be unique, one obtains

$$\partial g(\tau_1, \tau_2)/\partial t = -2\lambda K(t)g(\tau_1, t)g(t, \tau_2). \tag{59}$$

Since $g(\tau_1, \tau_2)$ is symmetric because of the symmetry of the kernel, (59) yields

$$dg(0, 0)/dt = -2\lambda K(t)g^2(0, t) \tag{60}$$

which proves (14c).

From (87), of Appendix II, follows[11]

$$\ln f = -\int_0^\lambda d\kappa\int_0^t d\tau K(\tau)g_\kappa(\tau, \tau) \tag{61}$$

where $g_k(\tau, \tau)$ is the solution of (58) with λ replaced by k. Differentiation with respect to t and use of (59) yields

$$d\ln f/dt = K(t)\int_0^\lambda [2\kappa g_\kappa^{(2)}(t, t) - g_\kappa(t, t)]\ d\kappa \tag{62}$$

where $g_k^{(2)}$ is defined by

$$g_\kappa^{(2)}(\tau_1, \tau_2) \equiv \int_0^t g_\kappa(\tau_1, \tau)K(\tau)g_\kappa(\tau, \tau_2)\ d\tau. \tag{63}$$

From (58) one obtains integral equations for $g_\kappa^{(2)}(\tau_1, \tau_2)$ and $\partial g_\kappa(\tau_1, \tau_2)/\partial\kappa$. Comparing these one finds

$$g_\kappa^{(2)}(\tau_1, \tau_2) = -\tfrac{1}{2}\partial g_\kappa(\tau_1, \tau_2)/\partial\kappa. \tag{64}$$

Substituting this result in (62) yields

$$d\ln f/dt = -\lambda K(t)g_\lambda(t, t) \tag{65}$$

and proves (14a).

The specific form of the kernel, $h(\tau_1, \tau_2) = e^{-\beta|\tau_1-\tau_2|}$ is needed for the verification of (14b) and (14d). The functions $g(\tau_1, \tau_2)$ and $e^{-\beta|\tau_1-\tau_2|}$ have a discontinuous derivative at $\tau_1 = \tau_2$, but we can obtain $dg(t, t)/dt$ from

$$\begin{aligned} dg(t, t)/dt = dg^*(t, t)/dt &= [\partial g^*(\tau_1, \tau_2)/\partial t]_{\tau_1=\tau_2=t} \\ &+ [\partial g^*(\tau_1, t)/\partial\tau_1]_{\tau_1=t} + [\partial g^*(t, \tau_2)/\partial\tau_2]_{\tau_2=t} \end{aligned} \tag{66}$$

where g^* is defined by

$$g^*(\tau_1, \tau_2) = g(\tau_1, \tau_2) - e^{-\beta|\tau_1-\tau_2|}. \tag{67}$$

The first term in (66) is evaluated by means of (59). The second term contains $[\partial g(\tau_1, t)/\partial\tau_1]_{\tau_1=t}$ which can be evaluated by differentiation of (58); one obtains

$$[\partial g(\tau_1, t)/\partial\tau_1]_{\tau_1=t} = \beta[2 - g(t, t)]. \tag{68}$$

The third term of (66) is equal to the second term by symmetry. Subtracting the corresponding differential quotients of $e^{-\beta|\tau_1-\tau_2|}$ one obtains (14b) with $\sigma_1 = g(t, t)$.

The proof of (14d) starts from

$$dg(0, t)/dt = [\partial g(0, \tau)/\partial t]_{\tau=t} + [\partial g(0, \tau)/\partial\tau]_{\tau=t} \tag{69}$$

and follows very closely the preceding derivation.

[9] G. N. Watson, "A Treatise on the Theory of Bessel Functions," Cambridge University Press, Cambridge, Eng., 2nd ed., p. 498; 1952.

[10] A. J. F. Siegert, "A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena. II. Examples," The RAND Corp., Paper P-730, sec, 3; September 1955. In this paper, the function h in (3.8), p. 16, should be replaced by f. The function g in (3.21), p. 19, should be replaced by $g^* \equiv g - \rho$. The right-hand sides of (3.23), p. 19, and (3.24), p. 20, should have positive signs, and the right-hand side of (3.25), p. 20, should be $\beta[2 - g(t, t)]$.

[11] Eq. (61) is essentially the expression for the Fredholm determinant of a kernel expressed in terms of its Volterra reciprocal function. See E. T. Whittaker and G. N. Watson, "A Course of Modern Analysis," Cambridge University Press, Cambridge, Eng., sec. 11.21, example 2, and sec. 11.22; 1940.

Appendix I

Eq. (21) can be written in the form

$$u'' + 2\beta \frac{dt(x)}{dx} (u' - u) = 0. \tag{70}$$

he Wronskian $\omega(x)$ defined by

$$\omega(x) = u_2' u_1 - u_1' u_2$$

n thus be conviently computed as

$$\omega(x) = \omega(a) \exp \{2\beta[t(a) - t(x)]\}. \tag{71}$$

$u_1(x)$, $u_2(x)$ are two linearly-independent solutions of 0) then the initial condition for $u(x)$ is satisfied by

$$x) = [u_2'(x_0) - u_2(x_0)]u_1(x) - [u_1'(x_0) - u_1(x_0)]u_2(x) \tag{72}$$

nless both coefficients vanish. Using (70) once more ne obtains

$$u(x) = -\left(2\beta \frac{dt}{dx}\right)_{x_0}^{-1} \{u_2''(x_0)u_1(x) - u_1''(x_0)u_2(x)\} \tag{73}$$

nd

$$\begin{aligned} (x_0) &= -\left(2\beta \frac{dt}{dx}\right)_{x_0}^{-1} d\omega(x_0)/dx_0 \\ &= -\omega(a)\, e^{2\beta t(a)} \end{aligned} \tag{74}$$

$$\begin{aligned} &= [u(x_0)/u(x)]^{1/2} = [\omega(a)\, e^{2\beta t(a)} \\ &\quad \cdot K(0)/2\beta\{u_2''(x_0)u_1(x) - u_1''(x_0)u_2(x)\}]^{1/2}. \end{aligned} \tag{75}$$

Appendix II

To obtain (8) for the function $\hat{r}(\eta \mid \zeta, t, \lambda)$ defined by), we write the Gaussian random function $x(t)$ in the rm

$$x(t) = \sum_\nu c_\nu \sqrt{\lambda_\nu}\, \varphi_\nu(t) \tag{76}$$

here the numbers λ_ν and the functions $\phi_\nu(t)$ are defined the eigenvalues and eigenfunctions of the integral uation

$$\int_0^t \rho(\tau - \tau')K(\tau')\varphi_\nu(\tau')\, d\tau' = \lambda_\nu \varphi_\nu(\tau) \tag{77}$$

ith normalization

$$\int_0^t K(\tau)\varphi_\nu^2(\tau)\, d\tau = 1, \tag{78}$$

nd where the random variables c_ν are independent and aussian with $\langle c_\nu \rangle_{\text{av}} = 0$ and $\langle c_\nu^2 \rangle_{\text{av}} = 1$.

We then have

$$\begin{aligned} &\left\langle \exp\left(i[\eta x(0) + \zeta x(t)] - \lambda \int_0^t K(\tau)x^2(\tau)\, d\tau\right)\right\rangle_{AV} \\ &\quad = \int \exp\,(i \sum_\nu c_\nu \varphi_\nu - \lambda \sum_\nu \lambda_\nu c_\nu^2) \prod_\nu \exp\left(-\frac{c_\nu^2}{2}\right) \frac{dc_\nu}{\sqrt{2\pi}} \end{aligned} \tag{79}$$

with

$$\psi \equiv \sqrt{\lambda_\nu}\, (\eta\varphi_\nu(0) + \zeta\varphi_\nu(t)), \tag{80}$$

since

$$\begin{aligned} \int_0^t K(\tau)x^2(\tau)\, d\tau &= \sum_{\mu\nu} \sqrt{\lambda_\nu \lambda_\mu}\, c_\nu c_\mu \\ &\quad \cdot \int_0^t K(\tau)\varphi_\nu(\tau)\varphi_\mu(\tau)\, d\tau = \sum_\nu \lambda_\nu c_\mu^2. \end{aligned} \tag{81}$$

Evaluation of the integral (79) yields

$$\begin{aligned} \hat{r}(\eta \mid \zeta, t, \lambda) &= \prod_\nu (1 + 2\lambda\lambda_\nu)^{-1/2} \\ &\quad \cdot \exp\left(-\frac{1}{2} \sum_\nu \frac{\lambda_\nu[\eta\varphi_\nu(0) + \zeta\varphi_\nu(t)]^2}{(1 + 2\lambda\lambda_\nu)}\right). \end{aligned} \tag{82}$$

We define the function $g_\lambda(\tau_1, \tau_2)$ by

$$g_\lambda(\tau_1, \tau_2) = \sum \frac{\lambda_\nu \varphi_\nu(\tau_1)\varphi_\nu(\tau_2)}{1 + 2\lambda\lambda_\nu} \tag{83}$$

and obtain from (77) that

$$\begin{aligned} g_\lambda(\tau_1, \tau_2) &+ 2\lambda \int_0^t \rho(\tau_1 - \tau')K(\tau')g_\lambda(\tau', \tau_2) \\ &= \sum_\nu \lambda_\nu \varphi_\nu(\tau_1)\varphi_\nu(\tau_2) = \rho(\tau_1 - \tau_2). \end{aligned} \tag{84}$$

In terms of $g_\lambda(\tau_1, \tau_2)$ we can now express the exponent in (82):

$$\begin{aligned} &\sum_\nu \frac{\lambda_\nu[\eta\varphi_\nu(0) + \zeta\varphi_\nu(t)]^2}{1 + 2\lambda\lambda_\nu} \\ &\quad = \eta^2 g_\lambda(0, 0) + 2\eta\zeta g_\lambda(0, t) + \zeta^2 g_\lambda(t, t). \end{aligned} \tag{85}$$

The product is obtained by writing

$$\begin{aligned} \frac{\partial}{\partial\lambda} \ln \prod_\nu (1 + 2\lambda\lambda_\nu)^{-1/2} &= -\sum_\nu \frac{\lambda_\nu}{1 + 2\lambda\lambda_\nu} \\ &= -\int_0^t K(\tau)g_\lambda(\tau, \tau)\, d\tau. \end{aligned} \tag{86}$$

Since the product is equal to unity when $\lambda = 0$, we get

$$\prod_\nu (1 + 2\lambda\lambda_\nu)^{-1/2} = \exp\left(-\int_0^\lambda d\kappa \int_0^t K(\tau)g_\kappa(\tau, \tau)\, d\tau\right). \tag{87}$$

On the Capacity of A Noisy Continuous Channel*

SABURO MUROGA†

Summary—The capacity of a noisy continuous channel is discussed in both cases where the signal transmitted over the channel is expressible by a process with mutually independent random variables and where it is expressible by a Markov process. Unlike discrete channels, continuous channels impose certain restrictions on transmitter power in general. In the case of a continuous channel under disturbance of additive noise, a theorem on the capacity in terms of channel parameters is obtained and applied. Then in the general case of a Markov process a general procedure to calculate the capacity is shown.

Introduction

THE channel capacity may be regarded as the most important concept introduced by Shannon in the information theory.[1] The author has already discussed the capacity of a noisy discrete channel.[2,3] We will now discuss that of a noisy continuous channel.

In general, the type of restriction on continuous channels is fairly different from that on discrete channels, so we must treat it separately.

If a function of time $f(t)$, which is an input signal to the channel, is limited to the frequency band 0 to W cps, we have the following expansion:

$$f(t) = \sum_{-\infty}^{\infty} X_n \frac{\sin \pi(2Wt - n)}{\pi(2Wt - n)}, \tag{1}$$

where $X_n = f(n/2W)$. This is a theorem which Someya[4] and Shannon derived independently. With this sampling theorem, we need consider only the ordinates $f(n/2W)$ at a series of discrete points spaced $1/2W$ seconds apart. Intervals between two successive points in time will be called Nyquist intervals.

A Continuous Channel Expressed by a Process with Mutually Independent Variables

The Capacity of a Continuous Channel Expressed by a Process with Mutually Independent Variables

Now we assume that the sampled values of signal $f(n/2W)$ are transmitted independently of each other over the continuous channel. Let $P(x)$ be the probability density that the signal value at the input of the channel at a sampled time point is x and $P'(y)$ be the probability density that the value at the corresponding point of time at the output is y. Let $p_x(y)$ be the conditional probability density that the signal value y is received at the output of the continuous channel under disturbance of a stationary noise when the signal value x is transmitted. This specifies the statistical property of this channel and will be called the channel transition density function.

Now the transmission rate R for this channel is defined as

$$R = -\int P'(y) \log P'(y)\, dy + \iint P(x)p_x(y) \log p_x(y)\, dx\, dy \tag{2}$$

and its capacity as the maximum of R for $P(x)$ with $p_x(y)$ fixed. But we have the following relation between $P(x)$ and $P'(y)$:

$$\int P(x)p_x(y)\, dx = P'(y). \tag{3}$$

We will introduce an auxiliary function $X(y)$ as follows in order to make maximization of (2) easier than a direct method.

Theorem I: Assume that the *dissemination characteristic equation*:

$$\int p_x(y)X(y)\, dy = \int p_x(y) \log p_x(y)\, dy \tag{4}$$

has a solution $X(y)$. Then the *dissemination*:

$$H_x(y) = -\iint P(x)p_x(y) \log p_x(y)\, dx\, dy \tag{5}$$

is expressed simply in the following form:

$$H_x(y) = -\int P'(y)X(y)\, dy. \tag{6}$$

Integrate the Fredholm's integral equation of the first kind (4) with x after multiplying it by $P(x)$ and then, since its right side is just $-H_x(y)$, we have

$$H_x(y) = -\iint P(x)p_x(y)X(y)\, dy\, dx.$$

If the order of the integrations in the right side of the above is interchangeable, we have finally (6), using the relation (3). Then the transmission rate takes the following simpler form:

$$R = -\int P'(y) \log P'(y)\, dy + \int P'(y)X(y)\, dy. \tag{7}$$

* Manuscript received by the PGIT, July 11, 1956.

† Elec. Communication Lab., Nippon Telegraph and Telephone Public Corp., Tokyo, Japan.

[1] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," Univ. of Illinois Press, Chicago, Ill.; 1949.

[2] S. Muroga, "On the capacity of a discrete channel I," *J. Phys. Soc. Japan*, vol. 8, pp. 484–494; July/August, 1953.

[3] S. Muroga, "On the capacity of a discrete channel II," *J. Phys. Soc. Japan*, vol. 11, pp. 1109–1120; October, 1956.

[4] I. Someya derived this theorem in "Theory of Waveform Transmission," (in Japanese), Shukyosha Co., Tokyo, ch. 4; 1944.

t should be noted that $X(y)$ is a known function which an be determined only from $p_x(y)$.

Therefore, the capacity for this channel could be btained if $P(x)$ obtained from (3) corresponding to $'(y)$ for which (7) is maximized under the condition $P'(y)\,dy = 1$ does not take a negative value over the efined interval of x. In the case of a noisy continuous hannel, however, we may have to consider additional estrictive conditions, for example like fixing the average ransmitter power or the peak transmitter power at a onstant value and like restriction to keep the average alue of the differences between the signal values of the ransmitter and of the receiver within a certain range. et us express it in general as follows:

$$\int S(x, y, P(x), P'(y))\,dx\,dy = \text{constant}. \tag{8}$$

Maximization of R under this restrictive condition ay be done conveniently by the Lagrange's method, o we need maximize only the following:

$$\begin{aligned} = &-\int P'(y)\log P'(y)\,dy + \int P'(y)X(y)\,dy \\ &+ \int \mu P'(y)\,dy + \int \lambda S(x, y, P(x), P'(y))\,dx\,dy \\ &+ \int \tau(y)\{\int P(x)p_x(y)\,dx - P'(y)\}\,dy, \end{aligned} \tag{9}$$

here μ, λ and $\tau(y)$ are coefficients to be determined. irst the calculus of variation of (9) with $P'(y)$ gives

$$\begin{aligned} -\log P'(y) - 1 + X(y) + \mu& \\ + \lambda \int \frac{\partial S}{\partial P'(y)}\,dx - \tau(y) = 0& \end{aligned} \tag{10}$$

nd that of (9) with $P(x)$ gives

$$\lambda \int \frac{\partial S}{\partial P(x)}\,dy + \int \tau(y)p_x(y)\,dy = 0. \tag{11}$$

10) gives $P'(y)$, that is,

$$'(y) = \frac{\exp\left(X(y) + \lambda \int \frac{\partial S}{\partial P'(y)}\,dx - \tau(y)\right)}{\int \exp\left(X(y) + \lambda \int \frac{\partial S}{\partial P'(y)}\,dx - \tau(y)\right)dy}, \tag{12}$$

while μ is determined from $\int P'(y)\,dy = 1$. From (11) we have

$$\int \tau(y)p_x(y)\,dy = -\lambda \int \frac{\partial S}{\partial P(x)}\,dy. \tag{13}$$

Insertion of (12) into (7) gives

$$\begin{aligned} = \log \int \exp\left(X(y) + \lambda \int \frac{\partial S}{\partial P'(y)}\,dx - \tau(y)\right)dy& \\ - \frac{\int \exp\left(X(y) + \lambda \int \frac{\partial S}{\partial P'(y)}\,dx - \tau(y)\right)\left(\lambda \int \frac{\partial S}{\partial P'(y)}\,dx - \tau(y)\right)dy}{\int \exp\left(X(y) + \lambda \int \frac{\partial S}{\partial P'(y)}\,dx - \tau(y)\right)dy}&, \end{aligned} \tag{14}$$

where λ and $\tau(y)$ should be determined from (3), (8), and (13). If $P(x)$ which can be calculated from thus obtained $P'(y)$ satisfies the required condition to be non-negative over the interval of x, C of (14) gives the capacity itself.

The Fredholm's integral equation of the first kind (13) cannot be solved explicitly in general. However, under a certain condition we can solve it. Here we will apply the Kameda's method on it.[5] Let $\tau(y)$ and $p_x(y)$ be expanded as follows:

$$\tau(y) = c_0\varphi_0(y) + c_1\varphi_1(y) + \cdots, \tag{15}$$

where $\varphi_n(y)$'s ($n = 0, 1, 2, \cdots$) are a system of the normalized orthogonal functions, for example

$$\begin{cases} \varphi_0(x) = \dfrac{1}{\sqrt{\pi}} \exp\left(-\dfrac{x^2}{2}\right) \\ \varphi_1(x) = \dfrac{1}{\sqrt{\pi}} \exp\left(\dfrac{x^2}{2}\right) \dfrac{d}{dx}\left(-\dfrac{x^2}{2}\right) \\ \varphi_2(x) = \dfrac{1}{\sqrt{\pi}} \exp\left(\dfrac{x^2}{2}\right) \dfrac{d^2}{dx^2}\left(-\dfrac{x^2}{2}\right) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases} \tag{16}$$

while the kernel $p_x(y)$ is expanded as follows:

$$p_x(y) = \chi_0(x)\varphi_0(y) + \chi_1(x)\varphi_1(y) + \chi_2(x)\varphi_2(y) + \cdots, \tag{17}$$

where coefficients, $\chi_0(x)$, $\chi_1(x)$, $\cdots$ are known functions. From a property of the normalized orthogonal functions, insertion of (15) and (17) into (13) gives

$$-\lambda \int \frac{\partial S}{\partial P(x)}\,dy = \sum_{n=0}^{\infty} c_n\chi_n(x). \tag{18}$$

Then our problem is reduced to the determination of coefficients, c_0, c_1, c_2, $\cdots$ in an expansion of the left side of (18) with a system of functions $\chi_n(x)$ ($n = 0, 1, 2, \cdots$). However $\chi_0(x)$, $\chi_1(x)$, $\cdots$, $\cdots$ are not always orthogonal to each other, so we can not expand it in general. But as a special case, if they are linearly in-

[5] T. Kameda, "A general method for solving linear integral equations," *Proc. Phys. Math. Soc. Japan*, vol. 10, Part I, pp. 231–235; 1927; vol. 11, Part II, pp. 17–27; 1928; vol. 11, Part III, pp. 169–180; 1928.

dependent of each other, we can find a system of orthogonal functions

$$\begin{cases} \theta_0(x) = a_{00}\chi_0(x) \\ \theta_1(x) = a_{10}\chi_0(x) + a_{11}\chi_1(x) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases} \tag{19}$$

from the above system of functions. Therefore we can determine the coefficients c_0, c_1, $\cdots$ by comparing the right side of (18) with the expansion of the left side of (18) with $\chi_0(x)$, $\chi_1(x)$, $\cdots$ which can be obtained by (19) after the left side of (18) is expanded with $\theta_0(x)$, $\theta_1(x)$, $\cdots$.

The Capacity of a Continuous Channel Where an Average Transmitter Power is Limited to a Certain Constant Value and $p_x(y) = k(y - x)$

If the noise is additive to the signal and also independent of it, we have $p_x(y) = k(y - x)$. We encounter many practical communication systems which noise of the above type disturbs. Now we assume that average power of the transmitted signals is fixed at a certain value P. That is,

$$\int x^2 P(x)\, dx = P \tag{20}$$

will be used instead of (8). Then we have

$$\frac{\partial S}{\partial P'(y)} = 0 \quad \text{and} \quad \frac{\partial S}{\partial P(x)} = x^2 f(y), \tag{21}$$

where

$$\int f(y)\, dy = 1. \tag{22}$$

We have

$$P'(y) = \frac{\exp\,(X(y) - \tau(y))}{\int \exp\,(X(y) - \tau(y))\, dy} \tag{23}$$

from (12) and

$$x^2\lambda = -\int \tau(y) p_x(y)\, dy \tag{24}$$

from (13). The integration of it over x after multiplying it by $P(x)$ and interchanging the order of integrations gives

$$\lambda P = -\int \tau(y) P'(y)\, dy. \tag{25}$$

and (14) is reduced to

$$C = \log \int \exp\,(X(y) - \tau(y))\, dy - P\lambda. \tag{26}$$

With $p_x(y) = k(y - x)$ we get

$$-\int p_x(y) \log p_x(y)\, dy = H(n) \tag{27}$$

which shows its left side to be equal to a constant value $H(n)$. Then a solution for the dissemination characteristic equation is

$$X(y) = -H(n). \tag{28}$$

Corollary I: If the noise is additive to the signal and also independent of it, the solution $X(y)$ of the dissemination characteristic equation is equal to the noise entropy $H(n)$ with the minus sign.

Therefore we have

$$\begin{cases} p'(y) = \dfrac{\exp\,(-\tau(y))}{\int \exp\,(-\tau(y))\, dy} \\[2ex] C = -H(n) + \log \int \exp\,(-\tau(y))\, dy \\[1ex] \qquad + \dfrac{\int \tau(y) \cdot \exp\,(-\tau(y))\, dy}{\int \exp\,(\tau(y))\, dy}. \end{cases} \tag{29}$$

Now we must determine $\tau(y)$ from

$$-x^2\lambda = \int \tau(y) k(y - x)\, dy, \tag{30}$$

which corresponds to (24).[6] With change of the argument, it is reduced to

$$-\lambda x^2 = \int \tau(x + z) k(z)\, dz. \tag{31}$$

Let us expand $k(z)$ and $\tau(y)$ into infinite series with a system of the Hermite function:

$$k(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \sum_{m=0}^{\infty} \frac{a_m}{m!} H_m(z) \tag{32}$$

and

$$\tau(y) = \sum_{n=0}^{\infty} b_n H_n(y), \tag{33}$$

where

$$\begin{cases} H_m(z) = (-1)^m e^{z^2/2} \dfrac{d^m}{dz^m} (e^{-z^2/2}) \\[2ex] a_m = \int k(z) H_m(z)\, dz \\[2ex] b_n = \dfrac{1}{n!\sqrt{2\pi}} \int \tau(y) H_n(y)\, e^{-y^2/2}\, dy. \end{cases}$$

[6] A formal solution for (30) could be obtained by means of the Fourier integral with some technique. (*Cf.* C. Titchmarsh, "Introduction to the Theory of Fourier Integrals," Oxford Press, Cambridge, England, p. 314; 1937.) But here the author is highly indebted to Zen'iti Kiyasu whose suggestion on the following method has made the solution of (30) more elegant.

Particularly we have $a_0 = \int k(z)\,dz = 1$ since $H_0(z) = 1$, and $a_1 = \int zk(z)dz$ is an average value $\bar{n}$ of the noise and furthermore we have $a_2 = \int(z^2 - 1)\,k(z)dz = \overline{n^2} - 1$ where $\overline{n^2}$ is a mean square value of the noise. It should be noted here that $k(z)$ in a_m and $\tau(y)$ in b_n are not weighed with the same function. Then we have

$$\tau(x + z) = \sum_{k=0}^{\infty} \tau^{(k)}(z)\frac{x^k}{k!} = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} b_{n+k}\frac{(n+k)!}{n!}H_n(z)\frac{x^k}{k!}. \qquad (34)$$

The right side of (31) may be expanded as follows:

$$\int \tau(x + z)k(z)\,dz = \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} b_{m+k}\frac{(m+k)!}{m!}\cdot\frac{x^k}{k!}a_m. \qquad (35)$$

Comparison of it with the left side of (31) gives us a set of the equations,

$$\begin{cases} \displaystyle\sum_{m=0}^{\infty} a_m b_{m+k}\frac{(m+k)!}{m!} = 0 \qquad (k \neq 2) \\ \displaystyle\sum_{m=0}^{\infty} a_m b_{m+2}\frac{(m+2)!}{m!}\cdot\frac{1}{2!} = -\lambda \end{cases} \qquad (36)$$

whose set of solutions are

$$\begin{cases} b_0 = -\lambda(2a_1^2 - a_2) \\ b_1 = 2a_1\lambda \\ b_2 = -\lambda \\ b_3 = b_4 = \cdots = 0. \end{cases} \qquad (37)$$

Finally we have

$$\tau(y) = -\lambda[(2a_1^2 - a_2)H_0(y) - 2a_1H_1(y) + H_2(y)] = -\lambda[2a_1^2 - a_2 - 2a_1y + y^2 - 1]. \qquad (38)$$

Calculation of λ, $P'(y)$ and C with these gives the following theorem, setting $\gamma = -\lambda$.

Theorem 2: If the noise is independent of the signal and amplitude of the received signal is a linear sum of amplitudes of the noise and the transmitted signal, the capacity of this continuous channel with an average transmitter power P is expressed as follows:

$$C = -H(n) + \frac{1}{2}\log\left(\frac{\pi}{\gamma}\right) + \gamma[P + \overline{n^2} - (\bar{n})^2] \qquad \text{(entropy/freedom)} \qquad (39)$$

where $H(n)$ is the noise entropy for this channel. And the probability $P'(y)$ which gives this capacity value to the transmission rate is

$$P'(y) = \sqrt{\gamma/\pi}\,\exp\left(-\gamma(y - \bar{n})^2\right), \qquad (40)$$

where

$$\gamma = \tfrac{1}{2}(P + \overline{n^2} - 3(\bar{n})^2)^{-1} \qquad (41)$$

and, $\bar{n}$ is an average value of the noise and $\overline{n^2}$ an average of the squared value of the noise. Here validity of this theorem requires that $P(x)$ corresponding to the above $P'(y)$ is non-negative over the whole defined interval of x.

Eq. (39) can be written in bits per second as follows:

$$C = 2W\left[-H(n) + \frac{1}{2}\log_2\left(\frac{\pi}{\gamma}\right) + \gamma(P + \overline{n^2} - (\bar{n})^2)\log_2 e\right] \qquad \text{(bits/second)} \qquad (42)$$

where W is the upper bound of frequency (cps) of the channel and the noise entropy $H(n)$ should be calculated in bits per degree of freedom.

Shannon's famous theorem on the capacity of a continuous channel under the disturbance of white noise may be obtained as a special case from this general theorem as follows.

Corollary 2: The capacity of a channel of band W cps disturbed by white noise of power N when the average transmitter power is P is given by

$$C = W\log_2\frac{P + N}{N} \qquad \text{(bits/second).} \qquad (43)$$

Proof: From the conditional probability density of the white noise

$$p_x(y) = \frac{1}{\sqrt{2\pi N}}\exp\left(-\frac{(x - y)^2}{2N}\right).$$

We have

$$X(y) = -H(n) = -\log_2\sqrt{2\pi eN}.$$

We have $\bar{n} = 0$, $\overline{n^2} = N$ and $\gamma = \frac{1}{2}(P + N)^{-1}$. Finally,

$$\begin{cases} C = \dfrac{1}{2}\log_2\dfrac{P + N}{N} \\ P'(y) = \dfrac{1}{\sqrt{2\pi(P + N)}}\exp\left(-\dfrac{y^2}{2(P + N)}\right) \\ P(x) = \dfrac{1}{\sqrt{2\pi P}}\exp\left(-\dfrac{x^2}{2P}\right). \end{cases} \qquad (44)$$

The first equation of the above is expressed in bits per degree of freedom for simplicity of calculation, so mutiplication of it with $2W$ gives the statement to be proven.

Next we will discuss a case of the noise with Rayleigh distribution.

Corollary 3: The capacity of a continuous channel of band W cps and of an average transmitter power P under the disturbance of the noise of power N with the Rayleigh distribution is given by

$$C = W\left[\log_2 \frac{2\pi}{e^2}\left(\frac{P}{N} - 1\right) + \left(\frac{P+N}{P-N}\right)\log_2 e\right] \quad \text{bits/second,} \tag{45}$$

provided that $P \gg N$, where

$$P'(y) = \frac{1}{\sqrt{2\pi(P-N)}} \exp\left(-\frac{(y-\sqrt{N})^2}{2(P-N)}\right) \tag{46}$$

$$P(x) = \frac{1}{\sqrt{2\pi}(P-N)^{3/2}}(P - x\sqrt{N}) \cdot \exp\left(-\frac{(x-\sqrt{N})^2}{2(P-N)}\right). \tag{47}$$

Proof: The Rayleigh distribution[7] is

$$W(z)\,dz = \begin{cases} \dfrac{1}{\sqrt{N}} \exp(-z/\sqrt{N})\,dz & z \geq 0 \\ 0 & z < 0 \end{cases}$$

where $\bar{z} = \sqrt{N}$ and $\overline{(z - \sqrt{N})^2} = N$. Then we have $\bar{n} = \sqrt{N}$, $\overline{n^2} = 2N$ and $\gamma = 1/[2(P-N)]$. Therefore we get (46) as $P'(y)$. From

$$H(n) = \log_e \sqrt{N} + 1, \tag{48}$$

we have the capacity in entropy per degree of freedom. Multiplication of it with $2W$ gives us the capacity in entropy per second. The probability $P(x)$ of (47) corresponding to this $P'(y)$ is obtained after some complex calculation of the Fourier integral of (3). However $P(x)$ in (47) does not satisfy the requirement as the probability density for it, since it is negative for $x > P/\sqrt{N}$ though its absolute value may be very small. If $P \gg \sqrt{N}$, we may be able to ignore the small difference between the above result and a correct one to be obtained. Consequently we get the above Corollary 3.

A Continuous Channel Expressed by a Markov Process

A Markov Process in a General State Space

The generalized continuous channels where not only sample sequences of signals but also noises are expressed as Markov processes, could be found easily in many real cases. Before discussing it, we will consider the difference between the Markov processes with continuous variables and those with discrete variables.[8]

When a signal value x_n at a certain point of time depends on a set of signal values at v points of time in the past; $x_{n-1}, \cdots, x_{n-v}$, the sequence of random variables is called a v-dimensional Markov process and particularly for $v = 1$ a simple Markov process. However since the vector process with random variables $\{\hat{x}_n\}$ where $\hat{x}_n = (x_{n-v+1}, \cdots, x_n)$, has the property of a simple Markov process, the v-dimensional Markov process can be reduced to a simple one. In the case of discrete variables where each x_n takes N different values, a sequence of random vector variables $\{\hat{x}_n\}$ can be considered a simple Markov process where a random variable takes N^v values and therefore an idea of a vector variable may not particularly be needed in the transformed simple Markov process. Moreover there is no essential difference from Shannon's diagrams which express these processes, as we see in the Shannon, Weaver, and Muroga.[1,2] The reduction of multiple Markov processes into simple ones can be applied even to a case of continuous variables, but the situation for the spacial expression is not so easy with increase of dimension as the discrete case.

Let X be a space of points ξ and let $\mathfrak{F}_X$ be a Borel field of X sets. A function $p(\xi, A)$ of $\xi \epsilon X$ and $A\ \epsilon \mathfrak{F}_X$ is called a stochastic transition function if it has the following properties. Particularly when it specifies a property of a channel, it will be called a channel transition function. That is, 1) $p(\xi, A)$ for fixed ξ determines a probability measure in A; 2) $p(\xi, A)$ for fixed A determines a ξ function measurable with respect to the field $\mathfrak{F}_X$.

The transition probability after n Nyquist intervals can be calculated as follows:

$$\begin{cases} p^{(1)}(\xi, A) = p(\xi, A) \\ p^{(n+1)}(\xi, A) = \displaystyle\int_X p^{(n)}(\eta, A) p(\xi, d\eta). \end{cases} \tag{49}$$

And the probability that signal value is in A at the n-th sample point is given by

$$\int_X p^{(n-1)}(\xi, A) P(d\xi), \tag{50}$$

which is $P(A)$ for $n = 1$. If it is independent of n, the process is strictly stationary and $P(A)$ is called a stationary absolute probability distribution.

Under a certain condition (Doeblin), many properties of a Markov process in a general state space are obtained with mathematical rigor. Roughly speaking, the Doeblin's condition is uniformity in ξ on the smallness of $p(\xi, A)$ for small A. This imposes a rather weak restriction on a property of channels. But most of the models of channels which have physical significance may satisfy it.

Under Doeblin's condition,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} p^{(m)}(\xi, E) = q(\xi, E) \tag{51}$$

exists uniformly in ξ and E, where $E\ \epsilon \mathfrak{F}_X$. A set E is called a consequent set if, for some ξ_0, $p^{(n)}(\xi_0, E) = 1$ for all n and then E is called a consequent of ξ_0. A set which is a consequent of every one of its points is called an invari-

[7] J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," Rad. Lab. Ser., McGraw-Hill Book Co., New York, N.Y.; vol. 24, p. 53; 1950.

[8] J. L. Doob, "Stochastic Processes," John Wiley and Sons, New York, N.Y.; 1953.

nt set. Such a set is either empty or has a finite valued neasure $\varphi(E) \geq \boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon}$ is a positive number. Like a liscrete case, decomposition of X is possible. If F is a set or which

$$\lim_{n\to\infty} p^{(n)}(\xi, F) = 0, \qquad \xi \epsilon X,$$

he set F is called a transient set. Assume a decomposition of X into disjunct invariant sets $E_1, E_2, \cdots$ and a transient et $F = X - \cup_a E_a$. And if there is a probability measure π of sets $E\ \epsilon \mathfrak{F}_X$ corresponding to each E_a such that

$$\pi(E_a) = 1, \qquad \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} p^{(m)}(\xi, E) = {}_a\pi(E) \quad \text{and} \quad \xi \epsilon E_a,$$

hen the E_a's is called ergodic sets. Furthermore E_a may be decomposed into cyclically moving subsets.

For simplicity we assume that there is only a single rgodic set which does not contain any cyclically moving subsets. Then the limit in (51) exists in an ordinary sense ather than Cesaro's and defines a stationary absolute probability distribution ${}_a\pi(E)$ independent of ξ. This is determined from (50):

$$P(A) = \int_X p(\xi, A) P(d\xi).$$

The Capacity of a Continuous Channel Expressed by a Simple Markov Process

Assume that when two signal values at the transmitter are x and x' respectively at two successive points of time which are one Nyquist interval apart, y and y' appear at the receiver at the two corresponding points. And we will define as a *channel transition density function* the conditional probability density $p_{xxy}(y')$ that the received signal at the fourth point of time is y' when signal values at the first three points of time among the above four points are x, x', and y, respectively. The integration of it with y' is the channel transition function. This specifies a statistical property of this channel.

Noise which is essentially a Markov process may be encountered often. For example, if noise is independent of signal and additive to it, $p_{xx'y}(y')$ is a function of only $x - x'$ and $y - y'$; $p_{xx'y}(y') = k(x - x', y - y')$. A signal also can be a Markov process, for example, like a radio receiver with avc in which a received amplitude y at some point of time varies the amplification gain of the receiver so that the signal at the next point may be affected, bringing y' as a received amplitude. Now consider ergodic sets at both transmitter and receiver, in similar fashion to a discrete case.[3] If $P_x(x')$ be the transition probability density, when the amplitude at the first of two successive points of time at the transmitter is x, the amplitude at the second point is x', and it specifies a property of this ergodic set at the transmitter. Then the stationary absolute probability density, that x is taken at a certain point of time, can be determined from

$$\int P(x) P_x(x')\, dx = P(x'). \tag{52}$$

The similar relation must hold at the receiver, replacing simply x and x' by y and y', respectively.

$P(x, y)$ is the joint probability density that the amplitudes at the transmitter and the receiver are respectively x and y at a certain point of time. $P(y, y')$ is the joint probability density that the amplitudes at two successive points of time at the receiver are respectively y and y'. Then from the property of the channel transition density function we have

$$\iint P(x, y) P_x(x') p_{xx'y}(y')\, dx\, dx' = P(y, y') \tag{53}$$

and

$$\iint P(x, y) P_x(x') p_{xx'y}(y')\, dx\, dy = P(x', y'). \tag{54}$$

Now let us introduce an auxilliary function $X(y, y')$.

Theorem 3: A solution for the following characteristic equation

$$\iint X(y, y') P(x, y) p_{xx'y}(y')\, dy\, dy'$$
$$= \iint P(x, y) p_{xx'y}(y') \log p_{xx'y}(y')\, dy\, dy' \tag{55}$$

is $X(y, y')$. Then the dissemination

$$H_x(y) = -\iiiint P(x, y) P_x(x') p_{xx'y}(y') \cdot \log p_{xx'y}(y')\, dx\, dx'\, dy\, dy'$$

can be simplified as follows:

$$H_x(y) = -\iint X(y, y') P(y, y')\, dy\, dy'. \tag{56}$$

A proof is similar to that of Theorem I.

Now, to find a capacity for this channel we will try to get the maximum of the following transmission rate with average transmitter power fixed:

$$-\iint P(y, y') \log \frac{P(y, y')}{\int P(y, y')\, dy'}\, dy\, dy' + \iint X(y, y') P(y, y')\, dy\, dy'. \tag{57}$$

Take this average transmitter power as P, that is;

$$\iint (x^2 + x'^2) P(x) P_x(x')\, dx\, dx' = P \tag{58}$$

or

$$\iint (x^2 + x'^2) P_x(x') P(x, y)\, dx\, dx'\, dy = P. \tag{59}$$

For convenience, the latter expression will be used for the following calculations.

Maximization of the transmission rate (57) subject to a constraint could be done by Lagrange's method. That is, we need to maximize only the following:

$$\begin{aligned} U = &- \iint P(y, y') \log \frac{P(y, y')}{\int P(y; y')\, dy'}\, dy\, dy' \\ &+ \iint X(y, y')P(y, y')\, dy\, dy' \\ &+ \int \eta(y)\left[\int P(y, y')\, dy' - \int P(y', y)\, dy'\right] dy \\ &+ \iiiint \delta(x, x')[X(y, y')P(x, y)p_{xx'y}(y') \\ &- P(x, y)p_{xx'y}(y') \log p_{xx'y}(y')]\, dx\, dx'\, dy\, dy' \\ &+ \iint \nu(x', y')\left[\iint P(x, y)P_x(x')p_{xx'y}(y')\, dx\, dy \right. \\ &\left. - P(x', y')\right] dx'\, dy' \\ &+ \iint \mu(y, y')\left[\iint P(x, y)P_x(x')p_{xx'y}(y')\, dx\, dx' \right. \\ &\left. - P(y, y')\right] dy\, dy' \\ &+ \iint \tau(x)P_x(x')\, dx'\, dx + \gamma \iint P(y, y')\, dy\, dy' \\ &+ \iiint \lambda(x^2 + x'^2)P_x(x')P(x, y)\, dx\, dx'\, dy. \end{aligned} \tag{60}$$

The third term in the above expression may be derived from other relations but is included here for convenience of calculation. Also

$$\left\{\begin{aligned} &\int P_x(x')\, dx' = 1 \\ &\iint P(y, y')\, dy\, dy' = 1 \end{aligned}\right. \tag{61}$$

are included in the above, as properties of the probability functions. Calculating the variation of $P(y, y')$ with U gives

$$\begin{aligned} -\log \frac{P(y, y')}{\int P(y, y')\, dy'} &+ X(y, y') + \eta(y) - \eta(y') \\ &- \mu(y, y') + \gamma = 0 \end{aligned} \tag{62}$$

and with $X(y, y')$ gives

$$\delta(x, x') = -P_x(x'). \tag{63}$$

Also, calculus of variations of it with $P_x(x')$ and $P(x, y)$ give respectively

$$\begin{aligned} &\iint \nu(x', y')P(x, y)p_{xx'y}(y')\, dy\, dy' \\ &+ \iint \mu(y, y')P(x, y)p_{xx'y}(y')\, dy\, dy' + \tau(x) \\ &+ \int \lambda(x^2 + x'^2)P(x, y)\, dy = 0 \end{aligned} \tag{64}$$

and

$$\begin{aligned} &\iint \delta(x, x')\Big[X(y, y')p_{xx'y}(y') \\ &- p_{xx,y}(y') \log p_{xx'y}(y')\Big]\, dx'\, dy' \\ &+ \iint \nu(x', y')P_x(x')p_{xx'y}(y')\, dx'\, dy' - \nu(x, y) \\ &+ \iint \mu(y, y')P_x(x')p_{xx'y}(y')\, dx'\, dy' \\ &+ \int \lambda(x^2 + x'^2)P_x(x')\, dx' = 0. \end{aligned} \tag{65}$$

Comparison of integration of (62) multiplied by $P(y, y')$ over y and y', and that of (65) multiplied by $P(x, y)$ over x and y, gives $\gamma = -\lambda P - C$. Insertion of it into (62) leads

$$\begin{aligned} \frac{P(y, y')}{\int P(y, y')\, dy'} = \exp \{&-C + X(y, y') + \eta(y) \\ &- \eta(y') - \mu(y, y') - \lambda P\}, \end{aligned} \tag{66}$$

whose integration over y' gives

$$\begin{aligned} e^{-\eta(y)} \quad W \int e^{-\eta(y')} \\ \cdot \exp \{X(y, y') - \mu(y, y') - \lambda P\}\, dy', \end{aligned} \tag{67}$$

where $\exp(-C) = W$. This is a homogeneous integral equation. We must find eigenvalues of W for it to have nonzero solutions. If the smallest negative log W could be found among eigenvalues and then all of probability densities satisfy the required conditions, $-$ log W gives the capacity for this channel. When the kernel of (67) is of convolution type, $e^{-\eta(y)}$ can be shown to have a fairly simple form under a certain condition.

Set Fredholm's determinant, that is, the following equation which may now be called a *capacity equation*, to be zero:

$$\begin{aligned} D(W) = 1 - \frac{W}{1!} \int K(\xi_1, \xi_1)\, d\xi_1 \\ + \frac{W^2}{2!} \iint \begin{vmatrix} K(\xi_1, \xi_1) & K(\xi_1, \xi_2) \\ K(\xi_2, \xi_1) & K(\xi_2, \xi_2) \end{vmatrix} d\xi_1\, d\xi_2, \end{aligned} \tag{68}$$

where $K(y, y')$ is the kernel of (67). In general, the eigenvalues for W can be obtained from it. When W_0 is a solution for $D(W) = 0$, an eignfunction can be obtained as a solution, that is,

$$\exp(-\eta(y)) = D(y, y_0', W_0), \tag{69}$$

where

$$D(y, y'; W) = WK(y, y') - \frac{W^2}{1!}\int \begin{vmatrix} K(y, y') & K(y, \xi_1) \\ K(\xi_1, y') & K(\xi_1, \xi_1) \end{vmatrix} d\xi_1 + \frac{W^3}{2!}\iint \begin{vmatrix} K(y, y') & K(y, \xi_1) & K(y, \xi_2) \\ K(\xi_1, y') & K(\xi_1, \xi_1) & K(\xi_1, \xi_2) \\ K(\xi_2, y') & K(\xi_2, \xi_1) & K(\xi_2, \xi_2) \end{vmatrix} d\xi_1\, d\xi_2 + \cdots \tag{70}$$

and moreover y_0' is chosen so as not to make $D(y, y_0', W_0)$ zero identically. If W_0 is a m-tuple solution of $D(W) = 0$, we have m number of solutions and then the most general solution can be expressed as a linear combination of them.

Comparison of the integration of a product of $P_x(x')$ and (64) over y', and the integration of a product of $P(x, y)$ and (65) over y, leads us to

$$\tau(x) = -\int \nu(x, y)P(x, y)\, dy, \tag{71}$$

so $\tau(x)$ vanishes from (64).

Summarizing the above—

Theorem 4: Obtain $P(y, y')$ from the eigenvalues W_0 and the eigenfunction $D(y, y_0', W_0)$ of the homogeneous integral equation (67) from the capacity equation (68), and solve the following equations;

$$\iint \mu(y, y')P(x, y)p_{xx'y}(y')\, dy\, dy' = -\iint P(x, y)p_{xx'y}(y')\{\nu(x', y') - \nu(x, y)\}\, dy\, dy' - \lambda\int (x^2 + x'^2)P(x, y)\, dy \tag{72}$$

$$\iint P_x(x')[p_{xx'y}(y')\log p_{xx'y}(y') - X(y, y')p_{xx'y}(y')]\, dx'\, dy' + \iint \nu(x', y')P_x(x')p_{xx'y}(y')\, dx'\, dy' - \nu(x, y) + \iint \mu(y, y')P_x(x')p_{xx'y}(y')\, dx'\, dy' + \int \lambda(x^2 + x'^2)P_x(x')\, dx' = 0 \tag{73}$$

$$P(y, y') = \iint P(x, y)P_x(x')p_{xx'y}(y')\, dx\, dx' \tag{74}$$

$$P(x', y') = \iint P(x, y)P_x(x')p_{xx'y}(y')\, dx\, dy \tag{75}$$

$$\iint X(y, y')P(x, y)p_{xx'y}(y')\, dy\, dy' = \iint P(x, y)p_{xx'y}(y')\log p_{xx'y}(y')\, dy\, dy'. \tag{76}$$

Then, $-\log W_0$ gives the capacity for this noisy channel for the smallest value of the eigenvalue W_0, for which the probabilities density from the above five equations satisfy the requirement for ergodicity.

Acknowledgment

The author is very grateful to Zen'iti Kiyasu, Chief of his section, for his advice in refining the material herein, and to Asst. Prof. Namio Honda for the stimulating inspiration he provided in initiating this work.

The author is also very grateful to Prof. Robert M. Fano for giving him a deeper appreciation of information theory at the Massachusetts Institute of Technology.

Merit Criteria for Communication Systems*

A. HAUPTSCHEIN† AND L. S. SCHWARTZ†

Summary—Merit criteria are presented for the transmission of information through noisy channels. The operational problem is formulated in such a way as to permit the minimum communication cost required to transmit a message over a communication system with a specified degree of reliability for given noise and loading (traffic density) conditions to be determined. The better system transmits the information at less cost. Cost values are shown to be functions of the basic merit parameters power, bandwidth and time, and the operational loading conditions. The formulation is general and can be applied to the evaluation of any modulation or coding system. Comparative evaluation of systems should result in performance indices which would permit judicious choice of systems for use in various operational situations.

Introduction

THIS PAPER formulates cost criteria for evaluating communication systems according to their ability to transmit information through noise. The paper uses entropy concepts of information theory and, in addition, discusses cost factors in relation to message density in operational situations. The cost criteria are designed to show that the better communication system has a given performance for less cost or a better performance for the same cost.

Information theory, in assigning a numerical measure to the information content of messages, provides an excellent starting point for judging system performance. In a situation requiring the transmission of a given amount of information, the parameters, power (P), bandwidth (W), and time (T), are interrelated through the expression of information rate. Moreover, information rate may be expressed in terms of per-unit equivocation which is a measure of reliability, since it expresses the residual uncertainty in the received information. Thus, a relationship exists between the P, W, and T parameters and reliability. In a manner to be discussed below, cost relations can be formulated such that the total cost of communication can be made a function of power, bandwidth, and time.

In evaluating a communication system the constraints which exist between P, W, and T for that system are known. The constraints permit a cost relation to be formulated, so that the values of the parameters which make the cost a minimum can be found. Having found the optimum values of the parameters for any one system of communication, one can then find the minimum cost of operating that system. In this manner, minimum cost for other systems can be determined and comparisons between systems made. That system which can transfer a given message with specified reliability at least cost is optimum.

Cost Formulation

In establishing communication cost relationships, one must include initial construction and installation costs, maintenance costs, and operational costs. Initial and maintenance costs are fixed costs and do not vary significantly with the operational situation in which they are used. On the other hand, operational costs, as defined here, are functions of the operational conditions. The fixed costs have been considered in engineering economy studies[1] and will not be treated here. It will be the main purpose of this paper to introduce the concept of operational cost and to show its effect on the optimization of communication systems. Communication cost will be defined as the sum of initial, maintenance, and operational costs.

A necessary first step in a cost formulation is to make a statement of preferences. In terms of cost, this means that relative weightings are assigned to the basic parameters of the communication problem. Relative weightings enter because, depending on the operational situation, the basic parameters may have different relative importance. Essential to the problem of cost estimation is the specification of the basic parameters which enter into the cost estimate. In communication systems, three parameters seem to be of fundamental importance and may be considered merit parameters. These are signal power, signal bandwidth, and operational time. These parameters are basic because: 1) They define information rate, 2) they determine initial and maintenance costs, and 3) they are essential to an operational cost formulation. Time, as used in this context, has two aspects. One aspect refers to the time to transmit a message over a channel; this is a function of the information, the information rate, and the coding delay. The other aspect refers to the waiting time and is a function of the percentage of full loading of the facilities and the number of communication channels. It is concerned with traffic density and problems of congestion such as arise in telephone operation and in aircraft control at large airports. Transmission of information results in information time delay and loading results in waiting time delay. The sum of the time delays is called the operational time delay.

Initial cost is a function of power, because power output determines the size and weight of a transmitter; it is a function of bandwidth, because bandwidth determines the size, character, and number of components; and it is a function of time, because time, when utilized in a tradeoff with power and bandwidth, implies the use of coders, decoders, and noise filters. Maintenance cost is a function

* Manuscript received by the PGIT, May 3, 1956. This work was sponsored under Contract No AF19(604)–1049.

† College of Eng., New York Univ., New York 53, N.Y.

[1] E. L. Grant, "Principles of Engineering Economy," The Ronald Press Co., New York, N.Y. 3rd ed.,; 1950.

all three merit parameters, because they determine quipment complexity.

Operational cost[2] is dependent on an ensemble of ossible future events. For example, power and time equirements determine energy demands which affect the eight and volume of equipment and, therefore, in the ase of airborne operations, the pay load. Also, if the aximum allowable time to transmit the location of a igh speed enemy aircraft is exceeded, the result may be atastrophic, and, consequently, the cost formulation ust reflect this fact.

To aid in further discussion, a set of channel cost gures are assigned $-{}_\nu C_P$, ${}_\nu C_W$, and ${}_\nu C_T$, called operational ower cost, bandwidth cost, and time cost respectively nder conditions of ν loading (density of traffic). In this iscussion the symbol ν refers to light (l), medium (m), r heavy (h) loading. It can, of course, refer to any degree f loading. The total operational cost ν^c on a per channel asis, under the various degrees of loading, can be expressed as the sum of the individual costs.

$$ {}_lC = {}_lC_P + {}_lC_W + {}_lC_T \quad (1) $$

$$ {}_mC = {}_mC_P + {}_mC_W + {}_mC_T \quad (2) $$

$$ {}_hC = {}_hC_P + {}_hC_W + {}_hC_T. \quad (3) $$

he total average operational cost can be obtained from e relation

$$ \bar{C}_0 = p(l)_lC + p(m)_mC + p(h)_hC \quad (4) $$

here $p(\nu)$ is the probability of the condition of ν loading. o this must be added initial and maintenance costs' hich are functions of the merit parameters P, W, and T ut independent of the degree of loading, and are symbolzed respectively as $C_I(P, W, T)$ and $C_M(P, W, T)$. The otal average communication cost $\bar{C}_c$ is then given by

$$ \bar{C}_c = \bar{C}_0 + C_I(P, W, T) + C_M(P, W, T). \quad (5) $$

The form of the cost relationships must be obtained om careful analysis of an operational situation. In eneral they may be represented by

$$ {}_\nu C_P = C_{P0}\, f(\alpha_\nu P_t) \quad (6) $$

$$ {}_\nu C_W = C_{W0}\, g(\beta_\nu W_t) \quad (7) $$

$$ {}_\nu C_T = C_{T0}\, h(\gamma_\nu T_0) \quad (8) $$

here P_t—average signal power.
W_t—signal bandwidth.
T_0—operational time; and is equal to the sum of information time T_i [transmission time plus encoding (T_e) and decoding (T_d) time] and waiting time T_W.
α_ν, β_ν, γ_ν—weighting constants which are determined from the operational situation.
C_{P0}, C_{W0}, C_{T0}—unit conversion or normalizing factors.

For reasons previously given, initial cost can be divided into component costs as

$$ C_I(P, W, T) = C_I(P_t) + C_I(W_t) + C_I(T_e + T_d). \quad (9) $$

Maintenance cost is probably a more complicated function of P, W, and T.

These cost equations must indicate, as realistically as possible, the exchange relationship existing between cost and the merit parameters. For example, cost must increase monotonically with increases in the merit parameters. The formulation should reflect a probable nonlinear rise of cost with P_t, W_t, and T_0, since there is usually some value of P_t, W_t, and T_0 beyond which it is extremely costly to operate because of equipment size, complexity, or available time. Weighting constants can be chosen to reflect this effect for different operational situations and degrees of loading.

The cost formulation should indicate that the merit parameters and unit increases in them are more costly for heavy loading than for light loading conditions. The reason for this is that for a given communication facility there is usually an upper limit to the available supply of power and the total assigned frequency band. For example, the total frequency band assigned for long wave communication may be in a frequency range 15 kc to 1500 kc. This is a fixed requirement that cannot be changed. If loading is light, meaning that relatively few channels in the band are used, individual channel bandwidths may be wide to overcome noise. Also, because of relatively small amount of traffic for the light loading condition, restrictions on operational time are not severe. On the other hand, if loading is heavy, meaning that all or nearly all of the available channels in the band are called in to use, an attempt to combat noise by use of wide bandwidth channels may be impermissable because of the limitation on over-all frequency band for the facility.

These statements suggest that for light loading, minimum power is more desirable than minimum bandwidth and minimum time. The implication of this statement is that for light loading wide bandwidth systems, redundancy coding, or integration should be used to combat noise. For heavy loading narrow bandwidth systems should be used because a premium is placed on bandwidth. Moreover, because waiting time and, therefore, operational time are increased in heavy loading, systems employing integration or lengthy coding techniques which increase information time should be avoided.

Considerations such as these must be employed in obtaining a cost formulation, and the process is one which rightly falls into the realm of Operations Research. The necessary values of parameters can be obtained from a study of past or current operations and they can be introduced into a theoretical model of the situation. This model, in our case, would be the proposed cost formulation. Once the model has been found to be a valid representation of the system under study then the values of the parameters may be altered in order to predict the effects of

[2] Since cost might refer to the dollars and cents value of an peration or to the manpower involved, the units of cost will have bitrary significance in this paper.

adopting new courses of action. This requires the accumulation of sufficient pertinent data.

Application of Merit Criteria

To illustrate some of the ideas just discussed, consider the problem of transmitting a message containing I bits of information in the presence of additive white Gaussian noise with a given reliability or per-unit equivocation. First, it is desired to determine the constraining relation among the merit parameters, and second, for an assumed operational situation, it is desired to determine the minimum cost condition.

Determination of Constraint Equation

It is assumed we have a message containing frequency components no greater than W_m and with significant amplitude for a length of time T_m. Outside this interval the amplitude is presumed small, so that most of the energy is confined within T_m and W_m. If

$$2T_mW_m \gg 1,$$

this message is determined to a high degree of accuracy by its values at $2T_mW_m$ sampling points spaced $\frac{1}{2}W_m$ seconds apart. If this message is quantized into B levels, it will contain at most

$$I = 2W_mT_m \log_2 B \text{ bits of information.} \tag{10}$$

Suppose it is desired to transmit this message over a symmetric binary (video) channel. This would require sending groups of $2^{\log_2 B}$ binary pulses. If n is the number of pulses per group, $2nW_m$ pulses are transmitted per second. This requires a signal bandwidth $W_t \geq nW_m$.

In terms of per unit equivocation E,[3] the channel information rate is

$$\begin{aligned} R' &= H(x) - H_y(x) \\ &= H(x)\left[1 - \frac{H_y(x)}{H(x)}\right] \\ &= 1 - E \text{ bits per symbol} \end{aligned} \tag{11}$$

where $H(x)$ = entropy before transmission, bits/symbol
= 1 for binary channel.

$$\begin{aligned} H_y(x) &= \text{equivocation, bits/symbol}^4 \\ &= -p \log p - (1-p) \log (1-p). \end{aligned} \tag{12}$$

$$\begin{aligned} p &= \text{probability of error} \\ &= \tfrac{1}{2}[1 - \Phi(\sqrt{(P_t/2\sigma^2)})] \end{aligned} \tag{13}$$

and

$$\begin{aligned} \Phi(a) &= \text{error integral} \\ &= \frac{2}{\sqrt{\pi}} \int_0^a e^{-x^2}\, dx. \end{aligned} \tag{14}$$

σ^2 = average noise power.

From (11) through (13) a relationship can be obtained between E and the merit parameter P_t.

As is often the case, this relationship is cumbersome to work with and an approximate expression is desirable. For binary pcm Jelonek[5] gives as an approximate expression for E

$$E = e^{-0.46P_t/\sigma^2} \tag{15}$$

or

$$P_t = -2.17\sigma^2 \ln E. \tag{16}$$

The rate at which information enters the channel per second can be determined from

$$\begin{aligned} R_b &= 2W_tH(x) \\ &= 2W_t \text{ bits per second.} \end{aligned} \tag{17}$$

The transmission time T_t is then obtained from (10) and (17) and is

$$T_t = \frac{I}{R_b} = \frac{W_mT_m \log_2 B}{W_t} \tag{18}$$

or

$$T_tW_t = W_mT_m \log_2 B. \tag{19}$$

In this example, encoding and decoding times[6] are assumed negligible compared to T_t, so that information time $T_i = T_t$. If, in addition, a condition of light loading exists and waiting time T_w is zero, $T_0 = T_t$.

Multiplying (16) and (19) determines a constraining relation for the merit parameters in terms of per-unit equivocation,

$$P_tT_0W_t = -2.17\sigma^2\, W_mT_m \log_2 B \ln E. \tag{20}$$

Cost Estimation

A prerequisite for a cost estimate is an operational situation. For illustrative purposes the following situation is assumed. Suppose there are N channels capable of operating from a total available power of P_T watts, and a total bandwidth of W_T cycles per second. Also, let T_T represent the total amount of operating time. Suppose that an analysis of similar past operating conditions indicates that an exponential relationship holds between the operational cost and the merit parameters. Then, (6), (7), and (8) become

$${}_\nu C_P = C_{P0}\, e^{\alpha_\nu P_t} \tag{21}$$

$${}_\nu C_W = C_{W0}\, e^{\beta_\nu W_t} \tag{22}$$

$${}_\nu C_T = C_{T0}\, e^{\gamma_\nu T_0}. \tag{23}$$

The weighting constants α_ν, β_ν and γ_ν, can be determined by specifying threshold percentages ${}_\nu x_p$, ${}_\nu x_w$ and ${}_\nu x_t$.

[3] R. M. Fano, "The Transmission of Information II," Res. Lab. of Electronics, M.I.T., Cambridge, Mass., Tech. Rep. No. 149; February, 1950.

[4] If the two levels of signal pulses are equiprobable.

[5] W. Jackson, "Communication Theory," Academic Press, New York, N.Y. p. 52; 1953.

[6] In this problem, by encoding time is meant the time to convert the original message into binary signal pulses; and vice versa for decoding time. In general, if coded systems are used, the time involved in coding the signal must be determined.

ere, ${}_\nu x_\mu$ represents that fraction of the total available nount of merit parameter (μ) above which the operaonal cost per unit increase in the merit parameters grows ; an excessive rate. From a careful analysis of the berational or tactical situation it should be possible to etermine with suitable accuracy the value of ${}_\nu x_\mu$. For cample the following values might be appropriate:

$$
\begin{aligned}
{}_1x_p &= 0.25; & {}_mx_p &= 0.10; & {}_hx_p &= 0.02\\
{}_1x_w &= 0.30; & {}_mx_w &= 0.10; & {}_hx_w &= 0.01\\
{}_1x_t &= 0.30; & {}_mx_t &= 0.10; & {}_hx_t &= 0.01.
\end{aligned}
$$

hese values reflect the relative importance of bandwidth nd time in heavy loading and that of power in light ading. To determine α_ν, β_ν, and γ_ν, the cost of ${}_\nu x_p P_T$ atts, ${}_\nu x_w W_T$ cycles per second and ${}_\nu x_t T_T$ seconds must e specified. If this cost is e^k, then

$$\alpha_\nu = \frac{k}{{}_\nu x_p P_T} \tag{24}$$

$$\beta_\nu = \frac{k}{{}_\nu x_w W_T} \tag{25}$$

$$\gamma_\nu = \frac{k}{{}_\nu x_t T_T}. \tag{26}$$

fact to note is that in the relative comparison of systems ie value of k may be removed by normalization. Also, r mathematical convenience, different degrees of loading n be simulated by suitably varying the information to e transmitted. This will maintain the form of the conraining relation (20).

For completion, further assume that from a study of ie equipment, the initial cost is found to be proportional P_t^2, W_t and $(T_e + T_d)^{1/2}$, and that maintenance cost proportional to (P_t, W_t, T_i). Substitution into $_I(P, W, T)$ and $C_M(P, W, T)$ gives the fixed cost assoated with the analysis.

ost Minimization

The total average cost $\bar{C}_c$ is obtained by substituting 24), (25), and (26) into (21), (22), and (23) and the sult into (5). It remains to determine the distribution the merit parameters which minimize this cost. This is adily done by employing Lagrange's method of unetermined multipliers. That is, form

$$(\overline{C_c} + \lambda P_t T_0 W_t) \tag{27}$$

where λ is the undetermined multiplier and then solve simultaneously (28), (29), (30), and (20).

$$\frac{\partial}{\partial P_t}(\overline{C_c} + \lambda P_t T_0 W_t) = 0 \tag{28}$$

$$\frac{\partial}{\partial W_t}(\overline{C_c} + \lambda P_t T_0 W_t) = 0 \tag{29}$$

$$\frac{\partial}{\partial T_0}(\overline{C_c} + \lambda P_t T_0 W_t) = 0 \tag{30}$$

where the constraint equation, (20) is

$$P_t T_0 W_t = -2.17\sigma^2 \, W_m T_m \log_2 B \ln E. \tag{20}$$

Following this procedure, the minimum cost behavior of a system with reliability E may be determined. In the comparison of systems for a fixed E, an equation like (20) must be determined for the different systems and the above minimization procedure carried out.

Conclusion

In this paper, merit criteria are presented for the transmission of information through noisy channels. The operational problem is formulated in such a way as to permit the minimum cost required to transmit a message over a communication system with specified reliability for given noise and loading conditions to be determined. Per-unit equivocation is used as a measure of reliability. The better system transmits the information at less cost. Cost is shown to be a function of the merit parameters, power, bandwidth, and time, and to vary with the operational situation. The formulation is general, requiring the determination of a constraining relation among the merit parameters only. Comparative evaluation of various modulation or coding schemes by the above method would result in performance indexes, permitting the judicious choise of systems for use in various operational situations.

Acknowledgment

The authors are grateful for many helpful suggestions from Dr. E. W. Sampson and C. F. Hobbs, of the Communication Laboratory, Air Force Cambridge Research Center and the Air Research and Development Command, Cambridge, Mass., and from F. Bloom of New York University, New York, N.Y.

First-Order Markov Process Representation of Binary Radar Data Sequences*

GEORGE C. SPONSLER†

Summary—Study of radar detection-trial data sequences has indicated the existence of interscan correlation. The theory of simple or first-order Markov chains is here applied to characterize the statistics of such sequences of correlated binary data consisting of detections (1's) and nondetections (0's) of a tracked target upon successive radar scans. Both stationary discrete and nonstationary continuous parameter processes are considered, for which relations are derived between four transition probabilities. $p_{i,j}$ and the absolute detection probability, β, and the so-called blip-scan ratio.

The discrete parameter, first-order Markov chain theory, presented first, is extended to the case wherein the blip-scan ratio may be expressed as a function of time. It is possible to employ the resulting nonstationary, continuous parameter solution to simulate radar data for aircraft flights of arbitrary patterns. Certain restrictions upon the admissible class of blip-scan functions are presented. In the case of the continuous parameter first-order process, the scan-to-scan correlation coefficient is shown to be restricted to positive values. An application is made to an automatic initiation problem.

Introduction

REPRESENTATION of radar data for purposes of theoretical study, simulation, or to aid in the design of automatic data-processing equipment, requires a mathematical model which will satisfactorily represent the statistics of such data. In this paper we shall develop such a representation for binary radar data, that is, for sequences of 1's and 0's representing detections and nondetections respectively, which are the output of some decision-making mechanism which decides whether or not a target is considered to be detected on each successive scan. The formal mathematical treatment of the problem in Section I is an application and extension of the theory of stationary, discrete parameter, two-state Markov chains; it is assumed *ab initio* that the radar statistics may indeed be represented by such first-order processes. In Section II, the theory is extended to a nonstationary, continuous parameter representation.

In the concepts to be developed, scan-to-scan, not pulse-to-pulse, relationships are considered, in which a search radar is presumed to be tracking a particular target. Each scan of the radar is a detection trial, in the statistical sense, resulting in either a detection or nondetection. This paper is concerned with the representation of the scan-to-scan, binary, detection-decision data of a search radar operating in a "track-while-scan" mode.

Study of radar data sequences at a number of different laboratories has indicated the existence of interscan correlation. To the author's knowledge, the possible existence of scan-to-scan correlation was first proposed in 1953 in a classified report by G. R. Lindsey of the Canadian Defence Research Board.[1] Lindsey's colleague, P. S. Olmstead of the Bell Telephone Laboratories, concurred in this proposal and subsequently developed a single-parameter statistical model employing a linearizing assumption which would represent radar binary data sequences for a particular type of aircraft and for a given interval of constant detection probability (*i.e.*, blip-scan ratio).[2] The work of the present author was initiated in an attempt to give a more precise mathematical basis for the description of scan-to-scan correlation and to extend the theory wherever possible.

Section I

It was the conclusion of Lindsey and Olmstead that binary radar data should be analyzed on the assumption that the probability of detection of an individual radar echo was explicitly dependent upon the detection (or nondetection) of the target echo of the immediately preceeding scan but upon no earlier scan. Thus, in the mathematical treatment, conditional detection probabilities were introduced.

As the detection probability upon each scan in this newer model depends in turn upon its predecessor, such probabilities are seen to form as interrelated chain of probabilities. In this section we assume the various probabilities are constant (such an assumption is satisfied by circular aircraft flights about the search radar). The statistics which relate the binary data of such a representation are an example of a stationary, discrete parameter, first-order, or "simple," two-state Markov chain, the two states being detection or nondetection (colloquially known as "hits" or "misses") respectively. The statistics are stationary as they are independent of the initial instant of time chosen, and they are discrete with sample period equal to the radar scan period.

The mathematical formulation of first-order, two-state Markov chains is well known and readily available in texts on probability theory.[3] A table of conditional probability elements is considered of the form:

* Manuscript received by the PGIT, July 12, 1956. The research in this document was supported jointly by the U. S. Army, Navy, and Air Force under contract with the Mass. Inst. of Tech.

† Lincoln Lab., M.I.T., Lexington, Mass.

[1] G. R. Lindsey, "Signal Correlation Between Successive Scans of a G.C.L. Radar," Defence Research Board–O.R.G.-34; March, 1953. (Title unclassified.)

[2] P. S. Olmstead, Bell Telephone Labs. Studies Case No. 27420, January 26, 1955, No. 27682, April 2, 1954.

[3] W. Feller, "Probability Theory and Its Applications," John Wiley and Sons, Inc., New York, N.Y., ch. 15; 1950.

$$[p_{i,j}] = \begin{array}{c|c|c} i \diagdown j & 0 & 1 \\ \hline 0 & p_{0,0} & p_{0,1} \\ \hline 1 & p_{1,0} & p_{1,1} \end{array} , \tag{1}$$

here the states 0 and 1 represent misses and hits, re-ectively, and $p_{i,j}$ is the conditional probability that a t or miss on any given scan will be followed by a hit or iss on the next scan. Thus, $p_{1,1}$ is the conditional robability that a hit will immediately follow a hit. he matrix $[p_{i,j}]$ of (1) is the stochastic matrix of a mple, two-state, Markov chain.

In order that the matrix (1) represent a physical radar etection model, it must be true that either a hit or a miss ill follow every hit or miss (*i.e.*, one or the other must ccur). Hence, it is necessary that

$$p_{0,0} + p_{0,1} = 1 \tag{2}$$

$$p_{1,0} + p_{1,1} = 1,$$

e., the sum of the conditional probabilities of the two ossible events following a hit or a miss must equal unity) in each case.

Furthermore, the total probability that a hit or miss ill follow an unknown preceding event must be equal β or $(1 - \beta)$ respectively, where β is the unconditional absolute detection probability, commonly known as e blip-scan ratio. β is thus the "single-look" probability at, knowing nothing of the preceding results, the signal eturn will be detected on a particular scan. The mathmatical statement of this definition is that

$$\left.\begin{aligned} (1 - \beta)p_{0,0} + \beta p_{1,0} &= (1 - \beta) \\ (1 - \beta)p_{0,1} + \beta p_{1,1} &= \beta \end{aligned}\right\} . \tag{3}$$

The last equation states, for example, that the unonditional probability of detection on a particular scan quals the sum of the probabilities that the preceding can-state, which is either a miss or hit, will undergo a ansition to a hit on the particular scan in question. For is reason, the $p_{i,j}$ are also often known as transition robabilities.

When one regards β as known, we thus have four quations, (2) and (3), in four unknowns (*i.e.*, $p_{0,0}$, ${}_{1,0}$, $p_{0,1}$, and $p_{1,1}$). Consideration of the system determiant, Δ, of this set of equations demonstrates that the ank of the linear set is 3 because $\Delta = 0$, but $\Delta = 0$ when ne equation is eliminated from the set. Hence, as shown the general theory of linear equations,[4] a one-parameter ystem of solutions exists. Mathematically, the parameter completely arbitrary; a convenient form of the solution given by

$$(p_{i,j}) = \begin{bmatrix} 1 - \alpha\beta & \alpha\beta \\ \alpha(1 - \beta) & 1 - \alpha(1 - \beta) \end{bmatrix}, \tag{4}$$

where α is the arbitrary parameter. This solution may be checked by direct insertion in (2) and (3). It is interesting to note the linear dependence of the $p_{i,j}$ upon β which justifies the linearizing assumption in Olmstead's model.

A more meaningful form of the solution is given by computing the scan-to-scan correlation coefficient, ρ, of a stationary, discrete parameter, first-order Markov chain. By definition the correlation coefficient between successive pairs of a random variable, x, is given by

$$\rho = \frac{\overline{(x_i)(x_{i+1})} - \bar{x}^2}{\sigma^2(x)} ,$$

where the bars indicate average values and x_i represents the value of x at time i. In our case all terms but one equal zero, leaving:

$$\rho = \frac{1 \cdot 1 \cdot \beta \cdot p_{1,1} - (1 \cdot \beta)^2}{(1 \cdot 1 \cdot \beta) - (1 \cdot \beta)^2} .$$

From (4) this expression becomes

$$\rho = \frac{\beta - \beta^2 - \alpha\beta(1 - \beta)}{\beta - \beta^2} ,$$

or

$$\rho = 1 - \alpha. \tag{5}$$

Thus the parameter α is equal to one minus the scan-to-scan correlation coefficient. Substituting for α in (4) we obtain the more meaningful solution:

$$[p_{i,j}] = \begin{bmatrix} \{(1 - \beta) + \rho\beta\} & \{\beta - \rho\beta\} \\ \{(1 - \beta) - \rho(1 - \beta)\} & \{\beta + \rho(1 - \beta)\} \end{bmatrix} . \tag{6}$$

If experimental data is available (preferably in large quantities) then an experimental stochastic matrix of the form given by (1) may be constructed. Let the experimental frequency of pairs of hits f_{11}, pairs of misses, f_{00}, and pairs of hits and misses, f_{01} and f_{10} respectively, be determined from the radar data, then it follows that

$$\begin{aligned} p_{0,0} &= \frac{f_{00}}{f_{00} + f_{01}} & p_{1,0} &= \frac{f_{10}}{f_{10} + {}_{11}} \\ p_{0,1} &= \frac{f_{01}}{f_{01} + f_{00}} & p_{1,1} &= \frac{f_{11}}{f_{11} + f_{10}} . \end{aligned} \tag{7}$$

Bartlett proves these expressions are actually the maximum likelihood estimators of the various transition probabilities of a stationary first-order Markov chain.[5] The statistics are ergodic so single-sequence and ensemble averages may be interchanged.

What now is the probability that n scans following an initial observation, the second observation will result in a hit or a miss, assuming nothing is known as to the

[4] F. B. Hildebrand, "Methods of Applied Mathematics," Pren-ce-Hall, New York, N.Y., Section 1.7 and 1.8, 1952.

[5] M. S. Bartlett, *Proc. Camb. Phil. Soc.*, vol. 47, pp. 86–95; 1951.

intervening events except their number? Such a question is answered in the theory of Markov chains by use of the higher-order transition matrix, $[p_{i,j}^{(n)}]$ (which is not to be confused with the stochastic matrix of a higher-order Markov chain). This transition matrix is the array of conditional or transition probabilities that n trials (scans) after an initial, observed event, the nth trial will result in a particular event (hit or miss). Then

$$[p_{i,j}^{(n)}] = \begin{array}{c|c|c} i \diagdown j & 0 & 1 \\ \hline 0 & p_{0,0}^{(n)} & p_{0,1}^{(n)} \\ \hline 1 & p_{1,0}^{(n)} & p_{1,1}^{(n)} \end{array} \qquad (8)$$

where $p_{i,j}^{(n)}$ represents the conditional probability that, if the initial observation on the first scan is represented by i, the second observation on the nth scan will be j.

The mathematical derivation of $[p_{i,j}^{(n)}]$ will not be presented here. The general result, derived by Feller, employing the stochastic matrix (6) gives:[6]

$$[p_{i,j}^{(n)}] = \begin{bmatrix} (1-\beta) & \beta \\ (1-\beta) & \beta \end{bmatrix} + \rho^n \begin{bmatrix} \beta & -\beta \\ -(1-\beta) & (1-\beta) \end{bmatrix}. \qquad (9)$$

It should be observed that if $n = 1$, then $[p_{i,j}^{(n-1)}] = [p_{i,j}]$ as would be expected. We should also expect that as the chain of events grows longer, the conditional probability elements of $[p_{i,j}^{(n)}]$ should approach the absolute probabilities, $(1 - \beta)$ and (β), in the respective columns. This means we would expect that as n grows larger, the particular initial event of the chain would have less and less influence on the result of the nth trial. If we let $n \to \infty$ in the expression (9) for $p_{i,j}^{(n)}$ we obtain the anticipated result, namely

$$\lim_{n\to\infty} [p_{i,j}^{(n)}] = \begin{bmatrix} (1-\beta) & \beta \\ (1-\beta) & \beta \end{bmatrix}, \qquad \rho \neq 1. \qquad (10)$$

This is the stochastic matrix of independent (*i.e.*, uncorrelated) events.

We now turn to the determination of the variance, σ^2, to be associated with the distribution of the number of detections in a sequence of scans generated by a simple Markov chain process. Let us define a stochastic variable N_n to be the number of hits in n scans. It is known that the distribution of N_n is asymptotically normal. The mean and variance thereof may be determined analytically from the mean and variance of another, related random variable distribution, namely, the distribution of the so-called recurrence times, X_k, to be defined hereafter. If μ and σ^2 are the mean and variance of this latter distribution, then the mean, $\bar{N}_n$, and variance, $\sigma^2(N_n)$, of the distribution of N_n are given by

$$\bar{N}_n \sim \frac{n}{\mu} \qquad (11)$$

$$\sigma^2(N_n) \sim \frac{n\sigma^2}{\mu^3}, \qquad (12)$$

as $n \to \infty$ the approximation becomes better.

We must, therefore, first determine the mean and variance of the distribution of the recurrence times.

It is shown in the theory of probability that, associated with an event, E, there is a random variable X_k, called the recurrence time, which is defined to be equal to the number of trials following the $(k - 1)st$ occurrence of E up to and including the kth occurrence.

Now under certain conditions satisfied in our problem, it is known that

$$\lim_{n\to\infty} p_{i,j}^{(n)} = \frac{1}{\mu_j} \qquad (13)$$

where $p_{i,j}^{(n)}$ is the i, jth element of the $[p_{i,j}^{(n)}]$ matrix, and μ_j is the mean value of the recurrence time of the state j.

We know from (10), for example, that $\lim_{n\to\infty} p_{1,1}^{(n)} = \beta$, and hence that $\mu_1 = 1/\beta$.

The explicit, analytical expression for $\sigma^2(X)$ is known to be,[7]

$$\sigma^2 = \mu_j - \mu_j^2 + 2\mu_j^2 \sum_{n=0}^{\infty} \left[p_{i,j}^{(n)} - \frac{1}{\mu_j} \right] \qquad (14)$$

if we are treating the recurrence time of an event the conditional probability of which is $p_{j,j}^{(n)}$, (note, we take $p_{j,j}^{(0)} = 1$). Let the event considered to be that of detection. From (9) we see

$$p_{1,1}^{(n)} = \beta + \rho^n \cdot (1 - \beta).$$

Hence

$$\sum_{n=0}^{\infty} \left[p_{1,1}^{(n)} - \frac{1}{\mu_1} \right] = \sum_{n=0}^{\infty} [\beta + \rho^n \cdot (1 - \beta) - \beta]$$

$$= \frac{1 - \beta}{1 - \rho}, \qquad | \rho | \neq 1.$$

Hence (14) becomes, (assuming hereafter $| \rho | < 1$),

$$\sigma^2(X) = \frac{1}{\beta} - \frac{1}{\beta^2} + \frac{2}{\beta^2} \frac{(1 - \beta)}{1 - \rho}. \qquad (15)$$

From (11) and (12) we have the estimates

$$\bar{N}_n \sim n\beta \qquad (16)$$

and

$$\sigma^2(N_n) \sim n\left[\beta^2 - \beta + \frac{2\beta(1 - \beta)}{1 - \rho} \right].$$

$$= n\beta(1 - \beta)\left[\frac{1 + \rho}{1 - \rho} \right], \qquad | \rho | \neq 1. \qquad (17)$$

It is interesting to observe at this point, that in the case of unconditional or independent events, $p_{1,1} = \beta$, (*i.e.*, $\rho = 0$) and (17) takes the form,

$$\sigma^2(N_n; \rho = 0) = n\beta\,(1 - \beta). \qquad (18)$$

[6] Feller, *op. cit.*, p. 351.

[7] Feller, *op. cit.*, p. 362.

his expression is merely the variance of the binomial stribution which is to be expected with independent vents, and hence provides an interesting check of the alidity of (17) in that case.

From the fact that $p_{1,1} = \beta + \rho(1 - \beta)$, we see that $_{,1}$ as a function of β will be given by a family of straight nes intersecting at the point (1, 1), with ρ the family arameter. It is known that $-1 \leq \rho \leq 1$. For positive alue of ρ, $(0 \leq \rho \leq 1)$, as all probabilities are limited to e interval zero to one, we see that $p_{1,1} \geq \beta$, which orresponds to the region above the 45° diagonal in Fig. The diagonal itself corresponds to the case of scan ndependence where $\rho = 0$. When ρ is negative, not all alues of $p_{1,1}$ are permissible. We know that $p_{1,1}$ and $_{,0}$, for example, must lie in the interval 0 to 1. Thus we equire,

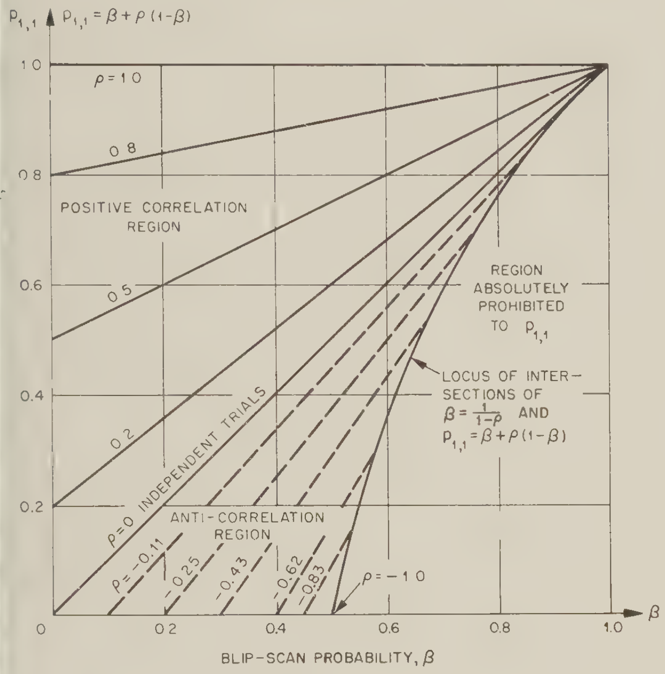


Fig. 1—Permitted values of $p_{1,1}$ for $-1 \leq \rho \leq 1$.

$$1 \geq \beta + \rho(1 - \beta) \geq 0$$

$$1 \geq \beta - \rho\beta \geq 0.$$

olving for β, we find

$$1 \geq \beta \geq \frac{-\rho}{1 - \rho} \tag{19}$$

$$\frac{1}{1 - \rho} \geq \beta \geq 0. \tag{20}$$

Now, if $0 \leq \rho \leq 1$, the inequalities (19) and (20) ogether require that $1 \geq \beta \geq 0$, of which we were aware o begin with. However, if $-1 \leq \rho \leq 0$, then we find

$$\frac{1}{1 - \rho} \geq \beta \geq \frac{-\rho}{1 - \rho}. \tag{21}$$

The right side of this inequality is always automatically satisfied in any plot of $p_{1,1}$ vs β. However, in the case of negative or anticorrelation, we must also satisfy the left side of inequality (21). Thus in the case of anticorrelation, $p_{1,1}$ as a function of β is restricted to the region below the 45° "independent-scan" diagonal bounded by thft diagonal, the axis $p_{1,1} = 0$, and the locus of the points of intersection of the various lines $p_{1,1} = \beta + \rho(1 - \beta)$ with the vertical lines $\beta = 1/(1 - \rho)$. This locus of intersections is plotted in Fig. 1, and the region of anticorrelation is depicted by the broken lines. (Note the degenerate point where $\rho = -1$ and $\beta = \frac{1}{2}$.)

Anticorrelation means a tendency exists toward alternation of hits and misses. Complete anticorrelation ($\rho = -1$) corresponds to sequences of the form (01010101 $\cdots$). To date, no such anticorrelation has been observed in actual radar data, and it is believed only the positive correlation region has physical realization. In this regard, Olmstead presents an experimental plot of $p_{1,1}$ (called x^px in his terminology) vs the blip-scan ratio, β, for a particular aircraft at a particular range and altitude. This graph is reproduced as Fig. 2; we note the linear dependence of $p_{1,1}$ on β and that we have positive correlation. Olmstead's model corresponds to a scan-to-scan correlation coefficient of $\frac{1}{2}$.

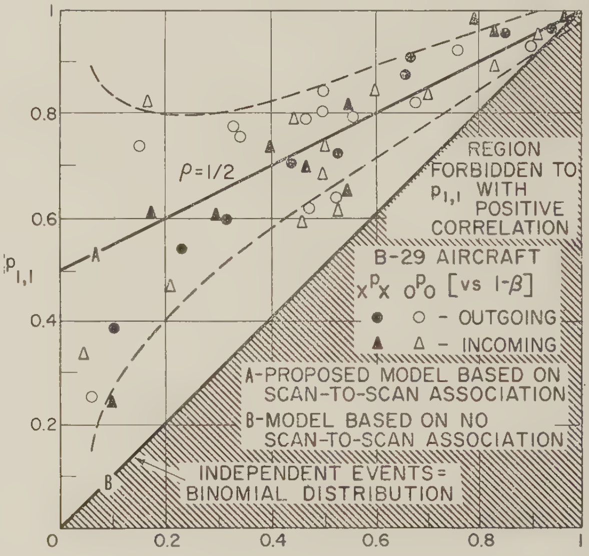


Fig. 2—$p_{1,1}$ vs β from Olmstead.

In this first section we have presented the statistics of a stationary, discrete parameter, first-order, two-state Markov chain process as applied to the representation of radar detection-trial data sequences. We have derived expressions for the transition probabilities and various related inequalities. We have also derived expressions for the mean and variance of the distribution of hits, N_n. We have not given any test to determine whether or not a simple Markov chain process is adequate to describe a particular set of radar data. Such a test, together with the statistics of higher order Markov processes are discussed along with other topics in a previous report by the author.[8]

Section II

In the previous section, the details of discrete parameter simple Markov chain binary data representation were developed on the assumption that the probabilities considered were not time (and hence not range) dependent. Strictly, that application of discrete parameter Markov chains is applicable only to airplane flights wherein the test aircraft flys a circular pattern about the radar. Thus, it is desirable to develop a Markov-type process which is applicable to nonradial (*i.e.*, arbitrary) flight patterns. Such a process would lead to markedly simplified radar data simulation, as well as to a more accurate and realistic representation of binary radar detection-trial data sequences.

The desired extension is obtained by use of nonstationary, continuous parameter, Markov processes. Of course, "nonstationary" means that the solution as a function of the time does *not* depend simply upon the difference in time which elapses between two successive looks (or between the two looks under consideration regardless of the number of scans coming between them), but rather is a function both of the initial and of the final observation times, s and t, the "continuous parameters". General expositions on such processes are to be found, among others, in the texts by Feller[9] and Doob.[10]

Analysis

In the mathematical treatment of continuous parameter Markov processes it is necessary to replace the fixed transition probabilities of the corresponding order discrete parameter Markov process with probability *functions*, that is by probabilities which are functions of the continuous parameter t, (*e.g.*, the time at which an aircraft under surveillance is at a particular range and azimuth). It is important to note in both the discrete and the continuous parameter cases we have been considering, that the aircraft is being tracked, *i.e.*, that it is under continuous surveillance. Therefore we are not attempting to *locate* an aircraft echo submerged in noise; rather, the radar is considered to "know" where the aircraft is at any particular time and decides (automatically or by observer) whether the result of a particular scan over the aircraft results in a detection [1] or nondetection [0], the data states.

The transition probability function of a nonstationary process must be a function of the two parameters, s, corresponding to the time at which the radar data is first observed, and t, corresponding to the later time at which time the data is observed for the second time. Thus the $p_{i,j}$ of the discrete parameter process are replaced by the function $p_{i,j}(s;\, t)$. Given that initially, at time, s, the statistics were in state i (= 1 or 0), the transition probability function $p_{i,j}\ (s;\, t)$ is the probability that upon the next "look" at the later time t, the radar data is in state j (= 1 or 0). It develops from the theory of continuous parameter Markov processes that these transition probability functions are solutions of a set of differential equations first derived by Kolmogorov. These differential equations themselves are directly obtainable, as will be shown, from the Chapman-Kolmogorov equation:

$$p_{i,k}(s;\, \mu) = \sum_j p_{i,j}(s;\, t) p_{j,k}(t;\, \mu), \qquad 0 \leq s < t < \mu. \tag{22}$$

This equation is a mathematical statement of the fact that the state k at time μ can be reached from the initial state i at time s through any intermediate state j at time t. Generally, it is impossible to solve for $p_{i,j}\ (s;\, t)$ directly from these equations; rather, a system of differential equations is derived from (22) from which the $p_{i,j}\ (s;\, t)$ may be determined provided certain boundary conditions are known.

Without going into the rigorous development, let us make the following definition of the functions[11] $c_i(t)$ and $c_{ij}(t)$ (compare Feller[12]).

Define

$$\left.\frac{\partial_i p_{,j}(s;\, t)}{\partial s}\right|_{s=t} = \left\{\begin{array}{ll} c_i(t), & i = j \\ -c_{ij}(t), & i \neq j \end{array}\right\} \tag{23}$$

$$\left.\frac{\partial p_{i,j}(s;\, t)}{\partial t}\right|_{t=s} = \left\{\begin{array}{ll} -c_i(s), & i = j \\ c_{ij}(s), & i \neq j \end{array}\right\}. \tag{24}$$

Now if in (22) we take partial derivatives with respect to s, thereafter setting s equal to t and then finally replacing the pair $(t;\, \mu)$ by the pair $(s;\, t)$ we find:

$$\frac{\partial p_{i,k}(s;\, t)}{\partial s} = c_i(s) p_{i,k}(s;\, t) - \sum_{j \neq i} c_{ij}(s) p_{j,k}(s;\, t). \tag{25}$$

Taking partial derivatives of (22) with respect to μ, and then setting μ equal to t, we obtain (5):

$$\frac{\partial p_{i,k}(s;\, t)}{\partial t} = -c_k(t) p_{i,k}(s;\, t) + \sum_{j \neq k} c_{jk}(t) p_{i,j}(s;\, t). \tag{26}$$

What appears to be single equations in both (25) and (26) are actually two systems of equations for all states i and k, where i and k, (having values 0 and 1 in the first-order case) could be vector states in a higher-order Markov process. The system of (25) is called the backwards system, because it involves differentiation with respect to the time of the first observation, s; the system (26) is called the forward system, because it involves differentiation with respect to the (later) time, t, of the

[8] G. C. Sponsler, "Markov Process Representation of Radar Detection-Trial Binary Data Sequences," Lincoln Lab. Tech. Rep. No. 114; April 16, 1956.

[9] W. Feller, "An Introduction to Probability Theory and Its Applications," John Wiley and Sons, Inc., New York, N.Y., ch. 17; 1950.

[10] J. L. Doob, "Stochastic Processes," John Wiley and Sons, Inc., ch. 6; 1953.

[11] Doob, *op. cit.*, p. 254.

[12] Feller, *op. cit.*, p. 388.

ond observation. In practical cases any solution of) will automatically be the solution of (26) and vice sa. Hence practically only one set of these equations d be solved to determine the $p_{i,j}\ (s;\, t)$.

n principle the Kolmogorov differential equations) or (26), permit the solution of the nonstationary, tinuous parameter Markov process of any order, vided one knows *a priori* the functions $c_i(t)$ and $c_{i,j}(t)$. later functions may be interpreted as probability sities; that is to say $c_k(t)$, for example, when multiplied a small increment of time, Δt, would approximate the bability that during the time interval t to $(t + \Delta t)$ a nge in state occurs out of state k. The actual probability ld include an error term, a function of t, which would roach zero Δt approached zero. Thus in practical cases, en treating higher-order, nonstationary, continuous ameter Markov processes, it is necessary to estimate $c_{i,j}(t)$ functions on the basis of such a probability erpretation. In most cases this functional dependence ar from obvious, and hence, although in principle a her-order nonstationary Markov process may be ved by use of (25) or (26), in practical cases it is quite icult because of lack of knowledge of the functional m of the $c_{i,j}$'s. Fortunately the first-order case turns to be comparatively simple in our radar application. e $c_{i,j}$'s in this case may be ascertained by a judicious essing process, and the validity of the results proven insertion in (22) through (26). We shall therefore monstrate the reasoning which leads to the recognition the solution for the nonstationary, continuous parameter, first-order Markov process.

n brief, what we shall do will be as follows: We shall [in (32) and (33)] that in the first-order case $c_{i,j} = c_i$. (36) and (37) we shall relate c_0 and c_1 to β, and then hibit a possible pair, c_0 and c_1, in (43) which 1) solve), and 2) reduce to a known form when β is constant. ce the reader accepts the plausibility of these c_i, he st accept (46) as an unique answer to the original blem. (This paragraph courtesy of the reviewer.)

Let us look then, at the Kolmogorov differential equans in the case of the nonstationary, continuous parameter, first-order Markov process. The subscripts i, j, k, ., take only the two values 0 and 1, and either (25) or) represents a system of four linear, differential equans in the unknowns $p_{0,0}(s;\, t)$, $p_{0,1}(s;\, t)$, $p_{1,0}(s;\, t)$, $_1(s;\, t)$. In our case it develops that the set of four uations (25) or (26), separate into two pairs of equations; h pair of which constitutes the differential equations a particular pair of $p_{i,j}(s;\, t)$. For example, the two ir of "forward" equations may be written in the form en by (27) and (28):

$$\frac{\partial p_{i,i}(s;\, t)}{\partial t} = -c_i(t)p_{i,i}(s;\, t) + c_j(t)p_{i,j}(s;\, t) \qquad (27)$$

$$\frac{\partial p_{i,j}(s;\, t)}{\partial t} = -c_j(t)p_{i,j}(s;\, t) + c_i(t)p_{i,i}(s;\, t) \qquad (28)$$

where $i \neq j$ and $i, j = 0, 1.$

Despite the appearance of the partial differential signs, these are actually ordinary differential equations. The equation for $p_{0,0}(s;\, t)$ would thus be written explicitly as

$$\frac{\partial p_{0,0}(s;\, t)}{\partial t} = -c_0(t)p_{0,0}(s;\, t) + c_1(t)p_{0,1}(s;\, t). \qquad (29)$$

The equation corresponding to (28), which would be associated with (29), would involve the derivative of $p_{0,1}(s;\, t)$.

In the preceding equations (27)–(29) we have tacitly assumed that $c_{ij}(t) = c_i(t)$. To prove this relation note that, corresponding to (2) of the previous section, we have the restraints:

$$p_{0,0}(s,\, t) + p_{0,1}(s,\, t) = 1 \qquad (30)$$

$$p_{1,0}(s,\, t) + p_{1,1}(s,\, t) = 1. \qquad (31)$$

If we differentiate both of these equations with respect to s, and then set $s = t$, we obtain:

$$c_0(t) = c_{01}(t) \qquad (32)$$

$$c_1(t) = c_{10}(t), \qquad \text{qed.} \qquad (33)$$

We shall next derive an important relation between $c_0(t)$ and $c_1(t)$. Let us assume that β is known as a function of t, For example, β might be a function of range or aspect or azimuth or all three; provided we know the aircraft trajectory we can always relate β to the single parameter, t. We recall that the $p_{i,j}(s;\, t)$ are the conditional probability functions which give the probability that at time t the system will be in state j after having been initially in state i at the earlier time s, no knowledge being assumed as to the intermediate states. Corresponding then to (3) of the previous section on Simple Markov Theory, we have for the continuous parameter case the analogous equations,

$$\beta(s)p_{1,1}(s;\, t) + [1 - \beta(s)]\cdot p_{0,1}(s;\, t) = \beta(t) \qquad (34)$$

$$\beta(s)p_{1,0}(s;\, t) + [1 - \beta(s)]\cdot p_{0,0}(s;\, t) = 1 - \beta(t). \qquad (35)$$

These two equations result from the definition of the blip-scan function; $\beta(s)$ and $\beta(t)$ are defined to be the single-scan or single-look absolute probabilities of detection at times s and t respectively. They are thus simply the values of the blip-scan function at times s and t.

Taking partial derivatives of either of (34) or (35) with respect to t, we obtain one and the same equation, namely:

$$-\beta(s)\ c_1(s) + [(1 - \beta(s)]\cdot c_0(s) = \beta'(s), \qquad (36)$$

where we have taken the limit as t approaches s after first having taken the derivatives. If we take partial derivatives of (34) and (35) with respect to s and then take the limit as s approaches t, we obtain the same equation, namely (36), with the parameter s replaced by t. This result is to be expected as the functions c_i are functions only of the instantaneous time. If we solve (36) for c_0, we obtain a general relation between c_0 and c_1 which must be satisfied, wherein we write the variable as τ, meaning that it may be any time:

$$c_0(\tau) = \frac{\beta'(\tau) + \beta(\tau)c_1(\tau)}{1 - \beta(\tau)}, \qquad \tau = s \text{ or } t. \tag{37}$$

In order to solve Kolmogorov differential equations, (27) and (28), it is first necessary to determine the form of the functions c_0 and c_1. As a guide in our solution, we know that for intervals of time separated by multiples of the radar scan period τ_s, the transition probability function for constant β must reduce to the higher-order transition matrix $[p_{i,j}^{(n)}]$ derived for the discrete first-order Markov process given by (9), namely

$$[p_{i,j}^{(n)}] = \begin{bmatrix} (1-\beta) & \beta \\ (1-\beta) & \beta \end{bmatrix} + \rho^n \begin{bmatrix} \beta & -\beta \\ -(1-\beta) & (1-\beta) \end{bmatrix}. \tag{38}$$

If an aircraft were flying a circumferential flight about the observing radar and if we started observing the aircraft at time s, then the exponent n in (38) would be approximated by

$$n = \frac{t - s}{\tau_s}, \tag{39}$$

where τ_s is the scan time of the radar. The equation is only approximate as the aircraft is moving; however, if

$$\frac{v_t \tau_s}{2\pi R} \ll 1, \tag{40}$$

then (39) is quite a good approximation. (Here v_t is the aircraft tangential velocity component and R is the aircraft range.)

If we replace n in (38) by the expression in (39) we would then expect to obtain the stochastic matrix of the continuous parameter process in the stationary case of constant β:

$$[p_{i,j}(s, t)] = \begin{bmatrix} (1-\beta) & \beta \\ (1-\beta) & \beta \end{bmatrix} + \rho^{(t-s)/\tau_s} \begin{bmatrix} \beta & -\beta \\ -(1-\beta) & (1-\beta) \end{bmatrix}. \tag{41}$$

Using (23) and (24), we may derive from (41) expressions for c_1 and c_0, the matrix of which would thus be given by

$$[c_{ij}(t)] = \frac{-\ln \rho}{\tau_s} \begin{bmatrix} \beta & \beta \\ (1-\beta) & (1-\beta) \end{bmatrix},$$

where we denote c_{00} by c_0 and c_{11} by c_1. Since in this case β is considered time independent, that is a constant, the derivative of β with respect to time is 0, and hence (37) is satisfied by these values of $c_{i,j}$. Hence, when β is a constant we surmise that

$$\left.\begin{aligned} c_0(t) &= \frac{-\ln \rho}{\tau_s}\beta \\ c_1(t) &= \frac{-\ln \rho}{\tau_s}(1-\beta) \end{aligned}\right\}. \tag{42}$$

It would be convenient if (42), valid for constant β, would also be applicable for the case where β is a nonconstant function of time. However such cannot be the case for in that event (37), which involves the derivative of β, would not be satisfied; however, (37) is satisfied if a single term in β' is added algebraically to (42) to give

$$\begin{aligned} c_0(t) &= \beta'(t) - \frac{\ln \rho}{\tau_s}\beta(t) \\ c_1(t) &= -\beta'(t) - \frac{\ln \rho}{\tau_s}[1 - \beta(t)]. \end{aligned} \tag{43}$$

Eqs. (43), (44), and (45) lead to an interesting observation. We have remarked the $c_i(t)$ function may be interpreted as probability densities from which fact we require $c_i(t) \geq 0$. Thus, from (43), (44), and (45) we require

$$\frac{\ln \rho}{\tau}\beta(t) \leq \beta'(t) \leq -\frac{\ln \rho}{\tau_s}[1 - \beta(t)]. \tag{44}$$

This inequality is a restriction upon the set of blip-scan functions admissible in our treatment of nonstationary, continuous parameter, first-order Markov processes.

As we require the $c_i(t)$ functions to be real, we also observe from (42) and (43) that the scan-to-scan correlation coefficient of our continuous parameter process is limited to positive values, *i.e.*,

$$0 \leq \rho \leq +1. \tag{45}$$

If this were not the case, then $\ln \rho$ would have an imaginary part. (Note incidentally that $-\ln \rho$ is positive for positive ρ.) Using (23), (43), (44), and (45) could be derived from the following transition probability function matrix, if (as we surmise) the general nonstationary, continuous parameter first-order Markov process solution were given by

$$[p_{i,j}(s, t)] = \begin{bmatrix} [1-\beta(t)] & \beta(t) \\ [1-\beta(t)] & \beta(t) \end{bmatrix} + \rho^{(t-s)/\tau_s} \begin{bmatrix} \beta(s) & -\beta(s) \\ -[1-\beta(s)] & [1-\beta(s)] \end{bmatrix}. \tag{46}$$

A simple calculation will demonstrate that (43), (44), and (45) are indeed thus derivable.

As may be proved by direct substitution, the solution of (46) satisfies both the Chapman-Kolmogorov equation (22), and the differential form of (27) and (28) with the $c_k(t)$ functions defined as in (23). It also satisfies the restrictions given by (30), (34), and (35), where $\beta(t)$ has been defined to be the first-order Markov process blip-scan probability function. The stochastic matrix of (46) is thus shown to be the required solution for the representation of binary radar data by means of a nonstationary continuous parameter, first-order Markov process.

In simulation applications where one desires to generate artifically sequences of binary data (a first-order process being assumed) one would employ the scan-to-scan form of (46) where τ is the radar scan period; *i.e.*,

$$_{i}(t - \tau, t)] = \begin{bmatrix} [1 - \beta(t)] & \beta(t) \\ [1 - \beta(t)] & \beta(t) \end{bmatrix} + \rho \begin{bmatrix} \beta(t - \tau) & -\beta(t - \tau) \\ -[1 - \beta(t - \tau)] & [1 - \beta(t - \tau)] \end{bmatrix}. \qquad (47)$$

is equation may be employed in practical cases to nulate sequences of binary data such as would be nerated by an aircraft flying any trajectory whatsoever, ovided that the blip-scan function is known as a funcn of time, that the statistics are indeed represented a first-order Markov process, and that the correlation constant.

Heuristically, (46) permits an easy interpretation of elf. The first term is the blip-scan probability (the tection probability) at the time, t, (that is, at the ne of the second observation on the radar data). This st term is modified by the correlated past history of the tistics according to the second term which is a function s, that is of the earlier time at which the first observan was made. The matrix (46), as we have seen in (43), tomatically reduces in the case of circular flights to the gher-order transition probability matrix of the discrete st-order Markov chain, if one measures time intervals tween successive "looks" in multiples of the radar an time.

It is interesting to note that if the correlation coefficient vere a function either s or t, the form of (43), (44), and 5) would not be changed. However, the derivitive of ρ th respect to s or t would appear in the differential uations (27) and (28), from which fact we note that ρ ay *not* be a function of s or t in a first-order Markov ocess according to our formulation. Thus, our present lution is limited to the case of constant scan-to-scan rrelation.

Another apparent restriction in generality imposed by r solution of (46) rises from the fact that the transition obabilities therein defined must lie in the closed interval to 1. In the case of $p_{1,1}(s; t)$, we require

$$1 \geq \beta(t) + \rho^{(t-s)/\tau_s}[1 - \beta(s)] \geq 0.$$

d from $p_{0,1}$ we require

$$1 \geq \beta(t) - \rho^{(t-s)/\tau_s}\beta(s) \geq 0.$$

The other two transition probabilities add no additional formation.) Together these inequalities require

$$- \rho^{(t-s)/\tau_s}[1 - \beta(s)] \geq \beta(t) \geq \rho^{(t-s)/\tau_s}\beta(s),$$
$$0 \leq \rho \leq 1. \qquad (48)$$

the scan-to-scan formulation of (47), this inequality comes

$$1 - \rho[1 - \beta(t)] \geq \beta(t + \tau) \geq \rho\beta(t). \qquad (49)$$

The inequality of (48) is only a reflection of the restricon of (44), as is easily proved by differentiating (48), d then letting s equal t.

Thus we are again reminded that not all possible blip-scan functions are admissible in our nonstationary, continuous parameter, first-order Markov process representation, but only those which satisfy the inequalities (44) and, equivalently, (48).

By way of an example, let us consider an application of the nonstationary, continuous parameter, first-order Markov process to the problem of automatic initiation. Let us assume that from a large number of aircraft test flights a blip-scan ratio, β, has been determined as a function of range and altitude for that particular aircraft. This determination might be secured from circular flights, for example, it being assumed that the circular flight β at a particular range is the same as the β for any aspect of the aircraft at that range. Let us further assume that this blip-scan ratio becomes vanishingly small at some given range (where, as defined later, $n = 0$). If an aircraft flies inbound, say along a radial course, what is the cumulative probability that it will be detected prior to range R? Our detection criterion will be that an automatic initiation facility exists whereby, when hits are weighted $+1$ and misses -1 and a cumulative sum is maintained, the aircraft will be considered to be detected whenever this cumulative sum reaches say sum-state 4 (extension to other automatic initiation schemes is obvious). Fig. 1 presents a ladder diagram of the possible sum-state transitions where the sum-states, k, are labeled 1, 2, 3, and 4, and the number of scans, n, are 1, 2, 3, 4, 5, and so on.

We assume the radar scan-to-scan detection probability is represented by a nonstationary, first-order, continuous parameter Markov process. Let $S(k: n)$ be the probability that the initiation sum is in state k, *exactly* n scans after $n = 0$. Let the cumulative detection probability, $C(4: N)$, be the probability that the aircraft will have been detected prior to or upon the Nth scan; then we see by our initiation criterion that $C(4: N)$ is given by

$$C(4: N) = \sum_{n=0}^{N} S(4: n). \qquad (50)$$

In the course of the analysis it develops that $S(k: n)$ enters only indirectly, rather what is required is the probability that the sum state k on the nth scan was the result of a transition in which either a hit or miss (that is scan-state 1 or 0) was the result of the nth scan. That is to say, we need the *joint* probability that sum-state k and scan-state i occur together on the nth scan. Let this joint probability be denoted by $S_i(k: n)$, wherein $i(=0 \text{ or } 1)$ is the scan-state resulting from the nth scan.

We observe that $S_1(0: n) = 0$ since it is impossible to enter sum-state 0 on any scan which results in a detection. Furthermore, as the aircraft flight is initiated, *i.e.*, detected, whenever $S(4: n)$ is obtained, (thereby halting the initiation process), we see that $S_0(3: n) = 0$ since no transition from sum-state 4 to 3 is possible. Thus we have six possible joint probabilities: $S_0(0: n)$, $S_0(1: n)$, $S_1(1: n)$, $S_0(2: n)$, $S_1(2: n)$, and $S_1(3: n)$; with

$$\left.\begin{aligned} S(0:n) &= S_0(0:n) \\ S(1:n) &= S_0(1:n) + S_1(1:n) \\ S(2:n) &= S_0(2:n) + S_1(2:n) \\ S(3:n) &= S_1(3:n) \end{aligned}\right\}. \qquad (51)$$

On Fig. 1, $S(k: n)$ are the sum-state probabilities at the point of intersection of the nth abscissa and the kth ordinate. The solid lines represent transitions resulting from hits, and the dashed lines represent transitions arising from misses. Consideration of Fig. 1 will show that:

$$\begin{aligned} S_0(0: n) &= S_0(1; n-1)p_{0,0}(n-1; n) \\ &\quad + S_1(1; n-1)p_{1,0}(n-1; n) \\ &\quad + S_0(0; n-1)p_{0,0}(n-1; n) \\ S_0(1; n) &= S_0(2; n-1)p_{0,0}(n-1; n) \\ &\quad + S_1(2; n-1)p_{1,0}(n-1; n) \\ S_1(1: n) &= S_0(0; n-1)p_{0,1}(n-1; n) \qquad (52) \\ S_0(2; n) &= S_1(3; n-1)p_{1,0}(n-1; n) \\ S_1(2; n) &= S_0(1; n-1)p_{0,1}(n-1; n) \\ &\quad + S_1(1; n-1)p_{1,1}(n-1; n) \\ S_1(3; n) &= S_0(2; n-1)p_{0,1}(n-1; n) \\ &\quad + S_1\ 2; n-1)p_{1,1}(n-1; n). \end{aligned}$$

Here the $p_{i,j}(n-1; n)$ are the scan-to-scan transition probabilities of the first-order, continuous-parameter Markov process given by (47) with t replaced by n and $\tau = 1$. Eqs. (53) are recurrence or difference equations; the initial conditions upon them are:

$$\begin{aligned} S_0(0, 1) &= 1 - \beta\ (1) \qquad (53) \\ S_1(1, 1) &= \beta(1) \\ S_0(1, 1) = S_0(2: 1) &= S_1(2: 1) = S_1(3: 1) = 0. \end{aligned}$$

Utilizing the recurrence (52) with the initial conditions (53), and employing the transition probabilities of (47), one may build up the probabilities of attaining various sum-states upon the various scans as indicated by the ladder diagram of Fig. 3. Such a difference equation method is particularly amenable to automatic computation as one need only store the results of the $(n-1)^{st}$ calculation in order to calculate the corresponding quantities on the nth scan. Thus using a computer to carry out the indicated calculation, the cumulative probability, $C(4: N)$, may be obtained. This solution is given in Fig. 4, where the resulting cumulative detection probability is given as a function of range from the initial scan, $n = 0$, with the sample blip-scan curve shown ($n = 0$ corresponds to a range of 202 miles). The various n's are those which would be obtained by an inbound aircraft at a speed which happens to make n equal to the aircraft range from the point $n = 0$. Cumulative detection probability curves for the automatic initiation criterion assumed are presented for the cases of scan-to-scan correlations of 0, 0.4, and 0.75. The limiting case, 1.0, is represented by the dashed line. As would be expected, the cumulative probability of detection improves as the (positive) scan-to-scan correlation grows larger.

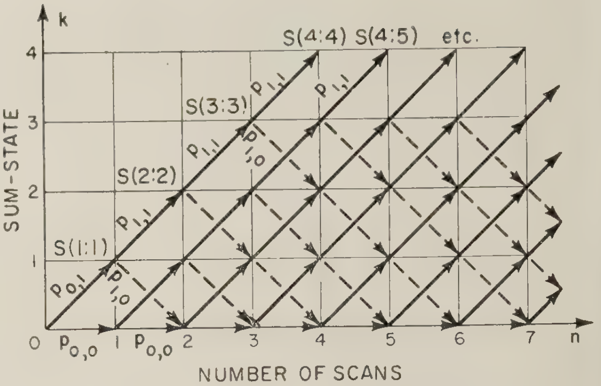


Fig. 3—Transition ladder diagram.

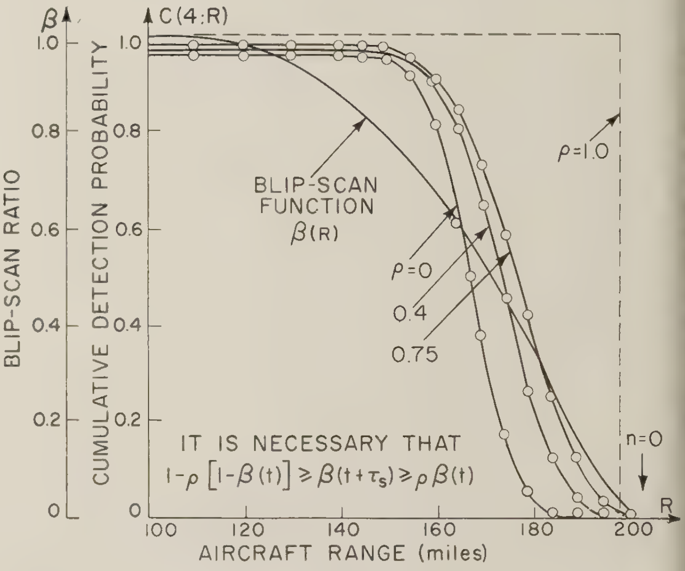


Fig. 4—Blip-scan function and cumulative detection probability.

Extension and further details of this study are available in the above mentioned technical report.[13]

Acknowledgment

The author wishes to acknowledge the aid of Dr. V. A. Nedzel, who initially suggested the study; Dr. E. J. Akutowicz and R. Gold, who proofread parts of the work; A. Ahlin and R. Holland, who carried out numerous numerical computations not herein included; and M. Carini, D. Sherrerd, and P. Restuccia, who aided in the preparation of the text.

[13] Sponsler, *op. cit.*

Automatic Bias Control for a Threshold Detector*

J. DUGUNDJI† AND E. ACKERLIND‡

ummary—A method for automatically controlling the thres- l bias in a detector is described and analyzed.

n Section I, the threshold bias problem is described: Bias is set a constant false alarm rate (or a constant false alarm time). 'standard biasing" is meant the common practice of adjusting required bias under the assumption that the noise is Gaussian has a flat power spectrum.

n Section II, the error that is made by standard biasing, if it s out the Gaussian noise does not have a flat power spectrum, ven.

n Section III, the automatic biasing method is given in the case re a constant false alarm time is required (its operation to ntain a constant false alarm rate is analogous). The device sioned operates as follows: One bias level $\tilde{v}_0$ is used as refer- e; the number of crossings per false alarm time T with positive e of the noise envelope through $\tilde{v}_0$ is averaged over a sufficiently time to yield a stable value C. This count C serves to determine threshold bias v. The level v changes only if the "long time" age count changes; it is specifically assumed that there is no onse to instantaneous changes in C. It is shown that such a ing method automatically adjusts to give a constant false m time (or rate), whenever the noise is Gaussian, and so has dvantage over the standard biasing method.

n Section IV, the efficiency of both methods for non-Gaussian e, is compared.

n Section V, the probability of detection of a short (relative he averaging time) "sure" signal with Rayleigh distributed plitude is given when automatic biasing is used; due to the plexity of the expression obtained, no direct comparison has n made with the case where standard biasing is used.

Section I

THE two "bias" problems in threshold detector may be formulated as follows: 1) Let an arbitrary band-pass and time $\tilde{T}$ be given. Determine the volt- level which the envelope of noise crosses, with positive pe, on the average of once per time $\tilde{T}$. 2) Let an arbi- ry band-pass and voltage level $\tilde{v}$ be given. Determine time T so that the envelope of noise crosses $\tilde{v}$ with itive slope on the average of once per time T.

Jnless explicitly stated otherwise, the noise considered e is Gaussian with zero mean, but not necessarily hite." The term "crossover" will always mean "cross- er with positive slope."

To solve either 1) or 2), one begins with Rice's equation.[1] e average number C of times per second that the noise velope crosses a level v (with positive slope!) is

$$C = \alpha(w)\cdot v\cdot \exp\left(-\frac{v^2}{2b_0}\right) \tag{1}$$

where

$$w = w(f) = \text{power spectrum of noise}$$

$$f_c = \text{midband frequency}$$

$$b_n = (2\pi)^n \int_0^\infty (f - f_c)^n w(f)\, df,$$

$$n = 0, 1, 2; \quad (b_0 = \text{mean noise power})$$

$$\alpha(w) = \frac{1}{b_0}\left[\frac{b_0 b_2 - b_1^2}{2\pi b_0}\right]^{1/2}.$$

Since TC is the average number of crossovers through v in time T, the voltage level and time in either of the problems 1) and 2) must satify the relation $TC = 1$, that is

$$1 = T\cdot\alpha(w)\cdot v\cdot \exp\left(-\frac{v^2}{2b_0}\right). \tag{2}$$

It is clear that this can be solved for T so long as $V\cdot\alpha(w) \neq 0$; however (2) cannot be solved for v unless T satisfies certain conditions.

A) In Appendix I it is shown that problem 1) has a solution $v(\tilde{T}, w)$ if and only if $\tilde{T}\cdot\alpha(w) \geq \sqrt{e/b_0}$, where e is the base of natural logarithms. Problem 2) always has the solution

$$T(\tilde{v}, w) = \frac{1}{\alpha(w)}\cdot\frac{1}{\tilde{v}}\cdot \exp\left[\frac{\tilde{v}^2}{2b_0}\right]$$

valid if $\tilde{v}\cdot\alpha(w) \neq 0$.

Whenever the condition $\tilde{T}\cdot\alpha(w) > \sqrt{e/b_0}$ is satisfied, (2) actually has two solutions $0 < v_1 < \sqrt{b_0} < v_2$ (see Appendix I); by elementary physical reasoning, the *largest* of these two solutions is the bias level desired in problem 1).

A flat power spectrum in the given band-pass is denoted by F, and one has

$$\alpha(F) = \beta\cdot\sqrt{\frac{\pi}{6b_0}} \tag{3}$$

where

$$\beta = \text{width of band-pass}$$

$$F(f) = b_0\beta^{-1},\ f \text{ in the band-pass}$$

$$= 0 \text{ otherwise.}$$

B) It follows from (3) and from the preceding section that, for white noise in the band-pass, problem 1) has a solution if and only if

$$\tilde{T} \geq \frac{1}{\beta}\sqrt{\frac{6e}{\pi}} \approx \frac{2.27}{\beta}.$$

* Manuscript received by the PGIT, October 17, 1956. The work cribed in this paper was done in connection with a Boeing Air- ft Co Subcontract.

† Dept. of Math., Univ. of Southern Calif., and consultant to dio Corp. of America, Los Angeles, Calif.

‡ Manager, Systems Engineering, RCA, Los Angeles, Calif.

[1] S. O. Rice, "Properties of a sine wave plus random noise," *l Sys. Tech. J.*, vol. 27, pp. 109–157; 1948. See (4.7)

Problem 2) always has the solution

$$T(\tilde{v}, F) = \frac{1}{\beta}\sqrt{\frac{6b_0}{\pi}}\cdot\frac{1}{\tilde{v}}\cdot\exp\left[\frac{\tilde{v}^2}{2b_0}\right].$$

Section II

In practice, especially when the power spectrum of the noise is subject to change during the course of operation of the detector, problems 1) and 2) are not solved by a continual recomputation of (1). In this part, the standard engineering practice, and its accuracy, will be considered. The problems 1) and 2) will be discussed separately.

Problem 1): The standard biasing method consists in measuring b_0, assuming that the noise has the flat power spectrum $F(f) = b_0\beta^{-1}$, and adjusting the threshold bias, once and for all, at the value $v(\tilde{T}, F)$. The error that can be committed by following this practice or, in other words, the range of variation in the average number of crossovers through $v(\tilde{T}, F)$, over all possible power spectra having the same mean power b_0, will now be determined. This requires two preliminary results.

A) Let $\tilde{T}$ and w_1 be such that $\alpha(w_1)\tilde{T} \geq \sqrt{e/b_0}$ so that there exists a solution $v_1 = v(\tilde{T}, w_1)$ for problem 1). Then for noise having power spectrum w and the same mean power as w_1, the average number of crossovers through $v(\tilde{T}, w_1)$ in time $\tilde{T}$ is

$$\frac{\alpha(w)}{\alpha(w_1)}$$

provided $\alpha(w_1) \neq 0$.

This is seen by noting that $\tilde{T}\alpha(w_1)v_1 \exp(-v_1^2/2b_0) = 1$ and that, according to (1), the average number of crossovers of the new noise through v_1 in time $\tilde{T}$ is

$$\tilde{T}\alpha(w)v_1 \exp\left(-\frac{v_1^2}{2b_0}\right) = \frac{\alpha(w)}{\alpha(w_1)}.$$

B) It is shown in Appendix II that $0 < \alpha(w)/\alpha(F) < \sqrt{3}$ for all power spectra in the given band-pass having the same mean power as the flat F. Values so close to 0 as desired are obtained by concentrating more power in any one frequency; values so close to $\sqrt{3}$ as desired are obtained by concentrating more power equally on both ends of the pass band.

With these preliminaries, one obtains immediately the possible error that can be made by following the engineering practice:

C) Let $\tilde{T} \geq 1/\beta\sqrt{6e/\pi}$ and assume standard biasing yields the level v_1. Then, regardless of how the noise is shaped, provided only that the mean power remains constant, the average number of crossings through v_1 in time $\tilde{T}$ will be between 0 and $\sqrt{3}$. Values as close to zero as desired will be obtained by concentrating a higher proportion of the power in a narrow portion of the band; values as close to $\sqrt{3}$ as desired are obtained by concentrating a higher proportion of the power equally at the ends of the band.

One merely applies A) and B) in this section, noting first that $\alpha(F)\tilde{T} \geq \sqrt{e/b_0}$.

Problem 2): The standard engineering method for solving 2) consists in measuring b_0 and taking

$$T(\tilde{v}, F) = \frac{1}{\beta}\sqrt{\frac{6b_0}{\pi}}\frac{1}{\tilde{v}}\exp\left[\frac{\tilde{v}^2}{2b_0}\right];$$

that is, in assuming that the noise has the flat power spectrum $F = b_0\beta^{-1}$ in the band. To compute again the error that can be made by following this practice, note that from the relation

$$T(\tilde{v}, F)\alpha(F)\tilde{v}\exp\left[-\frac{\tilde{v}^2}{2b_0}\right] = 1$$

$$= T(\tilde{v}, w)\cdot\alpha(w)\tilde{v}\exp\left(-\frac{\tilde{v}^2}{2b_0}\right)$$

one gets

$$T(\tilde{v}, w) = \frac{\alpha(F)}{\alpha(w)}T(\tilde{v}, F)$$

whenever $\alpha(w) \neq 0$. This, with II, B), leads at once to the following.

D) In problem 2), the range of variation in T necessary to maintain one crossover per time T at level $\tilde{v}$, over all changes in power spectrum having the same mean power and $\alpha(w) \neq 0$, is

$$\frac{1}{\sqrt{3}}T(\tilde{v}, F) < T < \infty.$$

Values as close as desired to the smallest value are obtained by concentrating a higher proportion of the power equally at both ends of the band; values as large as desired are obtained by concentrating more power in a narrower part of the band-pass.

Observe that

$$\frac{1}{\sqrt{3}}T(\tilde{v}, F) = \frac{1}{\beta}\left[\sqrt{\frac{2b_0}{\pi}}\frac{1}{\tilde{v}\exp\left(-\frac{\tilde{v}^2}{2b_0}\right)}\right].$$

From Appendix I, it follows readily thət, regardless of the level $\tilde{v}$ used, one always has

$$T > \frac{1}{\sqrt{3}}T(\tilde{v}, F) \geq \frac{1}{\beta}\sqrt{\frac{2e}{\pi}} \approx \frac{1.32}{\beta}.$$

Section III

In this part, an engineering practice (which will require somewhat more circuitry than that used in the standard method) for automatically solving problems 1) and 2) is described and discussed.

Select a voltage level v_0, which is to remain fixed once for all. Continually count the number of times per second that the noise envelope crosses v_0 with positive slope, and

$_0$ be a relatively long running average of these counts. currently also measure the mean noise power b_0. the C_0 and b_0 so obtained to calculate α from

$$C_0 = \alpha\cdot v_0\cdot \exp\left[-\frac{v_0^2}{2b_0}\right].$$

Problem 1): Whenever $\tilde{T}\cdot\alpha \geq \sqrt{e/b_0}$, the required bias level v is obtained by finding the largest root of

$$\tilde{T}\alpha\cdot v\cdot \exp\left[-\frac{v^2}{2b_0}\right] = 1.$$

Problem 2): Whenever $C_0 \neq 0$, the required time T is obtained from

$$T\alpha\tilde{v}\exp\left[-\frac{\tilde{v}^2}{2b_0}\right] = 1.$$

values of v or T so found change as C_0 and/or b_0 nce α) change.

s an alternative procedure, one can select two distinct age levels v_0, v_1, and obtain the numbers C_0, C_1 at se respective levels as before. The two simultaneous ations

$$C_0 = \alpha v_0 \exp\left[-\frac{v_0^2}{2b}\right]$$

$$C_1 = \alpha v_1 \exp\left[-\frac{v_1^2}{2b}\right]$$

e a unique solution (α, b_0) whenever $C_0 \neq C_1$; in this e one uses the (α, b_0) so obtained to proceed as before. The efficiency of this method for solving problem 1) hown here.

) Automatic biasing will adjust the bias level in blem 1) to maintain an average of one crossover time $\tilde{T}$, regardless of what the power spectrum and an power of the (Gaussian) noise is, provided only t $\tilde{T}\alpha(w) \geq \sqrt{e/b_0}$; that is, (see Section I, A) it will ust to the correct bias level whenever such a bias el exists. Indeed, the count C_0 obtained satisfies

$$C_0 = \alpha(w)v_0 \exp\left(-v_0^2/2b_0\right)$$

ere w is the actual power spectrum of the noise; the culated α is therefore exactly $\alpha(w)$, and the result ows.

3) In exactly similar fashion, automatic biasing will ust T in problem 2) to maintain an average of one ssover at $\tilde{v}$ per time T, regardless of the power spectrum l the mean noise power, provided only that $\alpha(w) \neq 0$; t is, it will adjust to the correct time interval whenever h a time exists.

These results apply, of course, also to the alternative cedure described above.

Section IV

The efficiency of the standard and automatic methods, in case the noise in the band-pass is non-Gaussian, is considered. Only problem 1) is discussed; a discussion of 2) can be done along identical lines.

Recall (Rice[1]) that the average number C' of times per second that the envelope of noise crosses the level v with positive slope is

$$C' = \int_0^\infty p(v, R')R'\,dR'$$

where $p(R, R')$ is the joint probability density of R and the slope R' at the same instant of time.

A) Let $\tilde{T} \geq 1/\beta\,\sqrt{6/\pi}$, and apply standard biasing to the non-Gaussian noise with probability density $p(R, R')$. The average number of crossovers C through $v = v(\tilde{T}, F)$ in time $\tilde{T}$ is

$$C = \frac{1}{\beta}\sqrt{\frac{6b_0}{\pi}}\cdot\frac{1}{v}\cdot\exp\left(\frac{v^2}{2b_0}\right)\cdot\int_0^\infty R'p(v, R')\,dR'.$$

For, by the hypothesis, a $v = v(\tilde{T}, F)$ exists satisfying

$$\alpha(F)\tilde{T}v\exp\left(-\frac{v^2}{2b_0}\right) = 1; \tag{4}$$

the expected number of crossovers by R through v in time T is

$$C = \tilde{T}\int_0^\infty R'p(v, R')\,dR'; \tag{5}$$

eliminating $\tilde{T}$ from (4) and (5), and using (3), gives the result.

B) Let automatic biasing be used, and assume that the experimental count C_0 at the fixed level v_0 satisfies

$$\tilde{T}C_0 \geq v_0\exp\left(-\frac{v_0^2}{2b_0}\right)\sqrt{\frac{e}{b_0}}$$

where b_0 is the experimentally measured mean noise power. Then the average number of crossovers C by the non-Gaussian R through the automatic bias level v in time $\tilde{T}$ is

$$C = \frac{v_0\exp\left(-\frac{v_0^2}{2b_0}\right)}{v\exp\left(-\frac{v^2}{2b_0}\right)}\cdot\frac{\int_0^\infty p(v, R')R'\,dR'}{\int_0^\infty p(v_0, R')R'\,dR'}.$$

In fact, the experimental count at v_0 will be

$$C_0 = \int_0^\infty p(v_0, R')R'\,dR'; \tag{6}$$

the α is determined from $C_0 = \alpha v_0\exp\left(-v_0^2/2b_0\right)$, and because of the condition on $\tilde{T}C_0$, the automatic bias level v exists: it satisfies

$$\tilde{T}\frac{C_0}{v_0}\exp\left(\frac{v_0^2}{2b_0}\right)\cdot v\exp\left(-\frac{v^2}{2b_0}\right) = 1. \tag{7}$$

The expected number of crossovers by R through v in time $\tilde{T}$ is

$$C = \tilde{T} \int_0^\infty R' p(v, R')\, dR'; \qquad (8)$$

eliminating $\tilde{T}$ from (8) and (7) and using (6) gives the result.

It should be observed that, unless $\tilde{T}C_0$ satisfies the condition stated, the automatic bias control does not yield any bias level; it may be possible that $\tilde{T} \geq 1/\beta \sqrt{6e/\pi}$, (so that standard biasing gives a level) even though $\tilde{T}C_0 < v_0 \exp(-v_0^2/2b_0) \sqrt{e/b_0}$ (so that automatic biasing gives no level); further discussion requires elaboration of $p(R, R')$. Note, however, that if one is using automatic biasing to solve problem 2), then the method yields a result whenever $C_0 \neq 0$.

Note, further, that if R, R' are independent, then whenever automatic bias control applies, the crossover rate is *independent* of the slope distribution, a state of affairs not true in standard biasing.

Section V

In this part, automatic and standard biasing will be compared according to their efficiency in detecting a sure signal in noise. The biasing will be according to problem 1).

Assume:

1) Automatic biasing is used.
2) A signal (sine wave superposed on the noise current) *will* appear during $0 \leq t \leq \mu\tilde{T}$ where $0 < \mu < 1$; its frequency is f_c.
3) The probability of the signal's appearance is uniformly distributed in $(0, \mu\tilde{T})$ independently of its amplitude. When it appears, it has duration Δ with zero buildup and decay times. Finally, its amplitude s has the Rayleigh distribution

$$\frac{s}{\sigma} \exp\left[-\frac{s^2}{2\sigma}\right].$$

The probability of detection (via a crossover at the bias level) is desired.

A) Assuming that the power spectrum $w(f)$ of the noise is symmetric about f_c, it is shown in Appendix III that the probability that the envelope passes through the automatic bias level v with positive slope during the time interval $(t, t + dt)$ is Ldt where

$$L = \frac{1}{\tilde{T}}\left(1 - \frac{\Delta}{\mu\tilde{T}}\right) + \frac{1}{\tilde{T}}\frac{\Delta}{\mu\tilde{T}}\frac{b_0}{b_0 + \sigma}\exp\left[\frac{1}{2}\frac{\sigma v^2}{b_0(b_0 + \sigma)}\right] + \frac{1}{\mu\tilde{T}}\int_0^\infty \gamma(s)\frac{s}{\sigma}\exp\left[-\frac{s^2}{2\sigma}\right] ds.$$

To define $\gamma(s)$, let $p(R \leq x)$ be the probability that the envelope of noise alone is at a voltage level $\leq x$. Then

$$\gamma(s) = \begin{cases} \frac{1}{2}p(v - s \leq R \leq v) - \frac{1}{\pi}\int_{v-s}^{v} \frac{R}{b_0}\exp\left[-\frac{R^2}{2b_0}\right] \cdot \arcsin\frac{v^2 - R^2 - s^2}{2sR}\, dR & [0 \leq s \leq v] \\ p(R \leq s - v) + \frac{1}{2}p(s - v \leq R \leq v) - \frac{1}{\pi}\int_{s-v}^{v}\frac{R}{b_0}\exp\left[-\frac{R^2}{2b_0}\right] \cdot \arcsin\frac{v^2 - R^2 - s^2}{2sR}\, dR & [v \leq s \leq 2v] \\ p(R \leq v) & [2v \leq s]. \end{cases}$$

B) As a consequence, with automatic biasing, the probability that a signal of strength s is acting at the time of a crossover at the bias level v is

$$\frac{1}{L}\frac{s}{\sigma}\exp\left[-\frac{s^2}{2\sigma}\right]\left\{\frac{\Delta}{\mu\tilde{T}}\frac{1}{\tilde{T}}\exp\left(-\frac{s^2}{2b_0}\right)I_0\left(\frac{sv}{b_0}\right) + \frac{1}{\mu\tilde{T}}\gamma(s)\right\}$$

where I_0 is the modified Bessel function of zero order. The probability that *no* signal is acting at the time of crossover is

$$\frac{1}{L}\frac{1}{\tilde{T}}\left[1 - \frac{\Delta}{\mu\tilde{T}}\right].$$

In this case that the standard biasing method is used, letting v_1 be the bias level, one has, instead of L, the expression

$$L_1 = \frac{\alpha(w)}{\alpha(F)}\left[\frac{1}{\tilde{T}}\left(1 - \frac{\Delta}{\mu\tilde{T}}\right) + \frac{1}{\tilde{T}}\frac{\Delta}{\mu\tilde{T}}\frac{b_0}{b_0 + \sigma} \cdot \exp\left\{\frac{1}{2}\frac{\sigma v_1^2}{b_0(b_0 + \sigma)}\right\} + \frac{1}{\mu\tilde{T}}\int_0^\infty \gamma_1(s)\frac{s}{\sigma}\exp\left(-\frac{s^2}{2\sigma}\right) ds\right]$$

where $\gamma_1(s)$ is the function described above, with v replaced by v_1 throughout.

Appendix I

First consider the function

$$\varphi(v) = \alpha T v \exp\left(-\frac{v^2}{2b_0}\right).$$

It is evident that $\varphi(v) = -\varphi(-v)$, that $\varphi(v) > 0$ for $v > 0$, that $\varphi(0) = 0$, and that $\varphi(v) \to 0$ as $v \to \infty$. The critical points occur whenever $d\varphi/dv = 0$, that is, at $\pm\sqrt{b_0}$. From the above, $v = \sqrt{b_0}$ is the unique maximum of φ; for $v \geq 0$, $\varphi(v)$ steadily rises till $v = \sqrt{b_0}$ and then steadily decreases.

A) The equation $\alpha T v \exp[-v^2/2b_0] = 1$ has:

1) Two solutions v_1, v_2 with $0 < v_1 < \sqrt{b_0} < v_2$ if $\alpha T > \sqrt{e/b_0}$ where e is the base of natural logarithms.
2) One solution $v = \sqrt{b_0}$ if $\alpha T = \sqrt{e/b_0}$.
3) No solutions if $\alpha T < \sqrt{e/b_0}$.

te first that the graph of $\varphi(v) - 1$ has the same form as t of $\varphi(v)$ but is depressed by one unit along the φ axis. solutions of $\varphi(v) = 1$ are the crossings of $\varphi(v) - 1$ with v axis. Now, if $\max \varphi(v) < 1$, the depressed curve will entirely below the v axis; if $\max \varphi(v) = 1$, the depressed ve touches the v axis at $v = \sqrt{b_0}$ alone; if $\max \varphi(v) > 1$, vill cross at a single point $0 < v_1 < \sqrt{b_0}$ and by the avior of φ as $v \to \infty$, also at a single point $v_2 > \sqrt{b_0}$. ce

$$\max \varphi(v) = \varphi(\sqrt{b_0}) = \alpha T \sqrt{b_0} \exp\left(-\tfrac{1}{2}\right)$$

result follows.

Appendix II

e proof is accomplished in three stages. Note that

$$\frac{\alpha(w)}{\alpha(F)} = \frac{\sqrt{3}}{\pi b_0 \beta} \sqrt{b_0 b_2 - b_1^2};$$

ce β, b_0 are constant, the quantity of interest is $B(w) =$ $_2 - b_1^2$.

Call "symmetric" power spectrum one satisfying $f_c - f) = w(f_c + f)$ where f_c is the midband. In the lowing, the graph of $w(f)$ is translated so that f_c curs at the origin.

A) For any given w, there exists a symmetric w_s having same mean power, with

$$B(w_s) \geq B(w).$$

rite

$$w(f) = \frac{w(f) + w(-f)}{2} + \frac{w(f) - w(-f)}{2}$$
$$\equiv w_s(f) + h(f).$$

is symmetric, and h is an odd function. Since $w_s \geq 0$ can be taken as a power spectrum. The mean power of is the same as that of w. In fact, because $h(f)$ is odd,

$$= \int_{-\beta/2}^{\beta/2} w_s(f)\,df = \int_{-\beta/2}^{\beta/2} [w_s(f) + h(f)]\,df$$
$$= \int_{-\beta/2}^{\beta/2} w(f)\,df = b_0;$$

cause w_s is even, $b_1' = 0$; finally

$$= 4\pi^2 \int_{-\beta/2}^{\beta/2} f^2 w_s(f)\,df$$
$$= 4\pi^2 \int_{-\beta/2}^{\beta/2} f^2[w_s(f) + h(f)]\,df = b_2.$$

us,

$$B(w_s) = b_0' b_2' = b_0 b_2 \geq b_0 b_2 - b_1^2 = B(w).$$

B) Let w be symmetric. Given any $\epsilon > 0$ there exists symmetric w_1 having the same mean power, such that vanishes everywhere in $-\beta/2 + \epsilon \leq f \leq \beta/2 - \epsilon$, d $B(w_1) \geq B(w)$.

Let

$$f_0 = \inf \{f \geq 0 \mid w(f) \neq 0\}.$$

The possibility $f_0 = 0$ is not excluded. Define $\eta = \frac{1}{2}(f_0 + \beta/2)$, set

$$u(f) = \begin{cases} 0 & 0 \leq f \leq \eta \\ w(f) + w(f + f_0 - \eta) & \eta \leq f \leq \dfrac{\beta}{2} \end{cases}$$

and extend $u(f)$ symmetrically over $-\beta/2 \leq f \leq 0$. The mean power of $u(f)$ is still b_0 since

$$\frac{b_0'}{2} = \int_0^{\beta/2} u(f)\,df = \int_\eta^{\beta/2} w(f)\,df$$
$$+ \int_\eta^{\beta/2} w(f + f_0 - \eta)\,df = \left(\int_\eta^{\beta/2} + \int_{f_0}^{\eta}\right) w(f)\,df = \frac{b_0}{2}.$$

By symmetry, $b_1' = 0$. Finally,

$$\tfrac{1}{2} b_2' = 4\pi^2 \int_\eta^{\beta/2} w(f) f^2\,df + 4\pi^2 \int_{f_0}^{\eta} w(f)[f + \eta - f_0]^2\,df.$$

Since for $f_0 \leq f \leq \eta$ one has $[f + \eta - f_0]^2 \geq f^2$ and since $w(f) \geq 0$, this shows

$$\tfrac{1}{2} b_2' \geq 4\pi^2 \int_{f_0}^{\beta/2} f^2 w(f)\,df = \frac{b_2}{2}.$$

Thus $B(u) \geq B(w)$ and $u(f)$ is certainly zero in $-\beta/4 \leq f \leq \beta/4$. Starting anew with $u(f)$ repeat the construction to obtain a $u_2(f)$ zero everywhere in $-(\beta/2^2 + \beta/2^3) \leq f \leq (\beta/2^2 + \beta/2^3)$ having mean power b_0, and satisfying $B(u_2) \geq B(u_1) \geq B(w)$. By repeating the construction N times, where

$$\sum_2^N \frac{\beta}{2^i} > \frac{\beta}{2} - \epsilon$$

the result follows.

Proof of Section II, B)

By A) and B) in this Appendix, $\sup B(w) = B(w_0)$ where

$$w_0 = \frac{b_0}{2}\left[\delta\left(f + \frac{\beta}{2}\right) + \delta\left(f - \frac{\beta}{2}\right)\right],$$

δ being the Dirac delta. A simple computation gives $B(w_0) = \pi^2 \beta^2 b_0^2$ so that $\alpha(w_0)/\alpha(F) = \sqrt{3}$. By elementary continuity considerations, there are power spectra w with $\alpha(w)$ so close to $\alpha(w)$ as desired, hence $\sup \alpha(w)/\alpha(F) = \sqrt{3}$.

On the other hand, one has $0 \leq B(w)$ and in fact, for $w_1 = b_0\,\delta\,(f - \varphi)$, φ any frequency in the band, actually $B(w_1) = 0$. Since again there are power spectra w with $\alpha(w)$ so close to $\alpha(w_1)$ as desired, this shows $\inf \alpha(w)/\alpha(F) = 0$.

Appendix III

Let: $r(t)$ = envelope of noise alone
$R(t)$ = envelope of noise plus signal s.

Fix a voltage level v_1 and assume known 1) the probability $c(v_1, s)dt$ that R crosses v_1 with positive slope in the time $(t, t + dt)$ and 2) the probability $\gamma(s)$ that there will be a crossover at the instant the signal s first appears. If an observation of the envelope be made, the probability that there will be an crossover during t, $t + dt$ will be determined.

Recall first that, if $p_G(H)$ is the probability of H, knowing G, the standard formula for conditional probabilities is

$$p(G)\cdot p_G(H) = p(G, H) = p(H)\cdot p_n(G) \qquad (9)$$

where $p(G, H)$ is the joint probability of G and H. Now let
G = crossover during $(t, t + dt)$,
H_1 = signal s present during entire time $(t, t + dt)$,
H_2 = signal s starts sometime during $(t, t + dt)$,
H_3 = no signal (*i.e.*, $s = 0$) during $(t, t + dt)$,
H_4 = signal ends in $(t, t + dt)$.
Using (9) gives

$$p_G(H_1) = kp(H_1)p_{H_1}(G) = kp(H_1)c(v_1, s)\,dt$$

$$p_G(H_2) = kp(H_2)\{\gamma(s) + A\,dt\}$$

$$p_G(H_3) = kp(H_3)c(v_1, 0)\,dt$$

$$p_G(H_4) = kp(H_4)E\,dt$$

where $k = 1/p(G)$ and A, E, are quantities readily calculated.

These four equations together determine one distribution. The k is therefore a normalizing factor; the sum of all the probabilities must equal 1, so that $p(G)$ can be determined from the condition

$$p_G(H_1) + p_G(H_2) + p_G(H_3) + p_G(H_4) = 1.$$

If s itself is a random variable, the normalizing equation is

$$\int p_G(H_1)p(s)\,ds + \int p_G(H_2)p(s)\,ds + p_G(H_3) + p_G(H_4) = 1. \qquad (10)$$

Turning to the problem in Section V one has

$$p\text{ (signal appears, at } t) = \frac{1}{\mu T}$$

and

$$p(H_1) = \frac{\Delta}{\mu T} - \frac{dt}{\mu T}$$

$$p(H_2) = \frac{dt}{\mu T}$$

$$p(H_3) = 1 - \frac{\Delta + dt}{\mu T}$$

$$p(H_4) = \frac{dt}{\mu T}.$$

Letting $p(s)$ denote the probability that the signal has amplitude s, one gets from (10), upon neglecting terms involving $(dt)^2$,

$$p(G) = \left(1 - \frac{\Delta}{\mu T}\right)c(v_1, 0) + \int_0^\infty \frac{\gamma(s)p(s)}{\mu T}\,ds + \int_0^\infty \frac{\Delta}{\mu T}c(v_1, s)p(s)\,ds. \qquad (11)$$

In the automatic biasing method, v_1 always satisfies

$$\alpha T v_1 \exp\left(-\frac{v_1^2}{2b_0}\right) = 1.$$

Since from Rice[2], using the assumption that $w(f)$ is symmetric around f_c,

$$c(v_1, s) = \alpha v_1 \exp\left[-\frac{v_1^2 + s^2}{2b_0}\right] I_0\left[\frac{sv_1}{b_0}\right],$$

one finds

$$c(v_1, s) = \frac{1}{T}\exp\left[-\frac{s^2}{2b_0}\right] I_0\left[\frac{sv_1}{b_0}\right]$$

$$c(v_1, 0) = \frac{1}{T}.$$

To calculate $\gamma(s)$ recall that

$$r = \sqrt{I_c^2 + I_3^2}, \qquad R = \sqrt{(I_c + s)^2 + I_s^2},$$

where I_c, I_s, are independent normal variates with zero means and variances b_0. Thus,

$$p(r \le \sqrt{I_c^2 + I_s^2} \le r + dr, \quad R \le \sqrt{(I_c + s)^2 + I_s^2} \le R + dR) \equiv p(r, R)\,dr\,dR$$
$$= \frac{1}{2\pi b_0}\iint_Q \exp\left(-\frac{x^2 + y^2}{2b_0}\right) dx\,dy$$

the region Q being obvious. This yields:

$$p(r, R) = \begin{cases} \dfrac{rR}{rb_0 s}\left[r^2 - \left(\dfrac{R^2 - r^2 - s^2}{2s}\right)^2\right]^{-\frac{1}{2}} \exp\left(-\dfrac{r^2}{2b_0}\right) & |\,r - s\,| \le R \le r + \\ 0 & \text{otherwise.} \end{cases}$$

The expression $\gamma(s) = p(r \le v_1, R \ge v_1)$ is easily computed from $p(r, R)$ and yields the expression given in the paper.

With evident calculations, one obtains the results in Section V, A) and B), from these expressions.

Acknowledgment

The authors wish to thank T. L. Gottier, Manager of West Coast Electronic Products Department, Radio Corporation of America, for his advice and encouragement.

[2] *Ibid.*, (4.8).

Exact Integral Equation Solutions and Synthesis for a Large Class of Optimum Time Variable Linear Filters*

JULIUS S. BENDAT†

ummary—This paper presents the exact integral equation ion and synthesis for a large class of optimum time variable r filters characterizing many physical problems. The signal om process is expressed in nonstationary Fourier series mble form, with certain statistical information assumed t its coefficients. The noise perturbation is represented by mped exponential-cosine autocorrelation function, which is of or importance in fields of physics and engineering, such as r, meteorology, and automatic control. For any finite operating od from 0 to *t*, the optimum time variable weighting function *t*) is found to be of a separable form, consisting of functions arameter τ multiplied by functions of parameter *t*, plus two a function contributions at the beginning and end. Valid hesis designs are developed for such separable weighting tions. Asymptotic synthesis techniques are formulated which r special situations of long-time or short-time operation. The lts are applied to two examples of practical interest.

Introduction

HE FIRST PORTION of this paper presents the solution of a general integral equation, occurring in prediction and filter theory, which gives the hematical form of the optimum time variable linear tem for a large class of practical problems. The input ssages $\{{}^j i(t)\}$ are assumed to be additive mixtures of als $\{{}^j s(t)\}$ and noise $\{{}^j n(t)\}$, where j, which may or y not be countable, denotes the different ensemble nbers of each random process. These messages are en to be zero for negative times so that their statistical perties are not invariant with respect to a shift in time. s corresponds physically to, say, having a starting tch in a system.

he input signals and the desired outputs $\{{}^j d(t)\}$ are mitted to be quite arbitrary in nature. The noise dom process is specified only through its autocortion function which is supposed to be of damped onential-cosine form. This type of noise occurs in ar fading records[1] and in wind gust velocities[2] to men- but two of many observed cases.

he desired output, ${}^j d(t)$, is any prescribed time ying linear operation on the signal component ${}^j s(t)$ ne; for example, ${}^j d(t)$ might be the signal ${}^j s(t)$ itself, its future value, or its integrated value, etc. The signal dom process is assumed to be a nonstationary random cess which is expressible in finite Fourier series ensemble form, with certain statistical information known about its coefficients. It is required to approximate ${}^j d(t)$ as closely as possible, for any finite time of operation from 0 to t, by means of a time varying linear filter acting on the full input message ${}^j i(t) = {}^j s(t) + {}^j n(t)$. The criterion used in defining the optimum system is to minimize the mean square ensemble difference between the actual output response of the system and the desired output.

The second portion of the paper considers the synthesis question. It is noted that the derived optimum time variable linear filter weighting functions $h(\tau, t)$ are of a separable form, consisting of functions of parameter t by functions of parameter τ, plus two delta function contributions at the beginning and end. A schematic design is developed for synthesizing such separable time variable weighting functions which is valid for all times of operation. For many problems, however, the explicit form of the optimum time variable weighting function may be so complicated that an exact mechanization for all times of operation is an extremely involved matter. On the other hand, the physical situation may indicate that the apparatus is to be used for a very short, or for a very long, period of time where in either of these asymptotic operating regions, the optimum weighting function assumes a simpler asymptotic form. Synthesizing an asymptotic weighting function instead of the optimum weighting function is shown to give an appropriate mechanization for the corresponding asymptotic region of operation.

In order to make the paper self-contained, Section I presents in condensed form the main ideas in the underlying prediction and filter theory. This theory originated from work by Wiener,[3] Zadeh-Ragazzini,[4] Booton,[5] and others. Section II takes up the general problem of interest in this paper and carries out the complete solution. The integral equation in question is shown in (16); the form of the answer appears in (21). Two applications are considered in Section III. A system design method for synthesizing separable time variable weighting functions of the mathematical form found in (21) is discussed in Section IV. Section V develops valid mechanizations for long-time operation for the two examples of Section III.

* Manuscript received by the PGIT, August 8, 1956.

† The Ramo-Wooldridge Corp., Los Angeles 45, Calif.

[1] H. M. James, N. B. Nichols, and R. S. Phillips, "Theory of vo-mechanisms," Rad. Lab. Ser., the McGraw-Hill Book Co., ., New York, N.Y., vol. 25, ch. 6; 1947.

[2] G. C. Clementson, "An Investigation of the Power Spectral hsity of Atmospheric Turbulence," Rep. No. 6445-T-31, (Sc. D. sis), Instrumentation Lab., Mass. Inst. Tech., Cambridge, ss.; May, 1950.

[3] N. Wiener, "The Interpolation, Extrapolation and Smoothing of Stationary Time Series," NDRC Rep., Cambridge, Mass., 1942; John Wiley and Sons, New York, N.Y.; 1949.

[4] L. Zadeh and J. Ragazzini, "An extension of Wiener's theory of prediction," *J. Appl. Phys.*, vol. 21, pp. 645–655; July, 1950.

[5] R. C. Booton, "An optimization theory for time-varying linear systems with nonstationary statistical inputs," M.I.T. Meteor Rep. No. 72; July, 1951, and Proc. IRE, vol. 40, pp. 977–981; August, 1952.

The Conclusion summarizes the principal results, while the Appendix contains details of mathematical proofs.

Section I. Basic Theory

Consider a possibly nonstationary collection of input messages $\{{}^ji(t)\}$ composed of a mixture, not necessarily additive, of input signals $\{{}^js(t)\}$ and perturbing noises $\{{}^jn(t)\}$, j denotes the different ensemble members. Suppose that this message passes for a finite operating period T through a particular time variable linear system characterized by its weighting function $h(\tau, t)$, where $h(\tau, t)$ denotes the response of the time varying system at time t to a unit impulse applied at time $t - \tau$. For physical realizability, it is necessary that $t - \tau \leq t$ or $\tau \geq 0$. The actual output response of the system is given by

$$ {}^jr(t) = \int_0^T h(\tau, t)\, {}^ji(t - \tau)\, d\tau. \qquad (1)$$

The above formula for ${}^jr(t)$ represents infinite operating time procedures by letting T approach infinity. Independent of operating time, it applies to constant parameter linear systems by replacing $h(\tau, t)$ by $h(\tau)$, where $h(\tau)$ measures the response of a constant parameter system to a unit impulse after a time τ has elapsed. For the important practical case of a starting switch in the circuit, or where the input message is zero for negative time, T becomes a variable t.

The desired output, ${}^jd(t)$, is permitted to be any time varying linear operation on the signal component of the message, ${}^js(t)$ itself. The difference between the actual output response ${}^jr(t)$ and the desired output ${}^jd(t)$ is the system error ${}^je(t)$,

$$ {}^je(t) = \int_0^T h(\tau, t)\, {}^ji(t - \tau)\, d\tau - {}^jd(t). \qquad (2)$$

It is required to determine the weighting function $h(\tau, t)$ such that the mean square ensemble average of ${}^je(t)$, namely $\langle {}^je^2(t)\rangle_{\text{Av over } j}$, is a minimum. Squaring and averaging ${}^je(t)$ over j gives,

$$\begin{aligned} \langle {}^je^2(t)\rangle_{\text{Av}} = \int_0^T \int_0^T & h(\nu, t)h(\tau, t) \\ & \cdot \langle {}^ji(t - \nu)\, {}^ji(t - \tau)\rangle_{\text{Av}}\, d\tau\, d\nu \qquad (3) \\ - 2\int_0^T & h(\nu, t)\, \langle {}^ji(t - \nu)\, {}^jd(t)\rangle_{\text{Av}}\, d\nu \\ & + \langle {}^jd(t)\, {}^jd(t)\rangle_{\text{Av}}. \end{aligned}$$

By definition, for nonstationary processes, the autocorrelation function of the input message with itself, or the desired output with itself, is

$$\gamma_{ii}(t_1, t_2) = \langle {}^ji(t_1)\, {}^ji(t_2)\rangle_{\text{Av over } j} \qquad (4)$$

$$\gamma_{dd}(t_1, t_2) = \langle {}^jd(t_1)\, {}^jd(t_2)\rangle_{\text{Av over } j} \qquad (5)$$

the angular bracket notation indicating ensemble average, the remaining notation showing dependence upon the times of observation. The cross-correlation function of the input message and the desired output is

$$\gamma_{id}(t_1, t_2) = \langle {}^ji(t_1)\, {}^jd(t_2)\rangle_{\text{Av over } j}. \qquad (6)$$

In terms of the correlation functions, the mean square ensemble system error shown in (3) becomes

$$\begin{aligned} \langle {}^je^2(t)\rangle_{\text{Av}} = & \int_0^T \int_0^T h(\nu, t)h(\tau, t)\gamma_{ii}(t - \nu, t - \tau)\, d\tau\, d\nu \\ & - 2\int_0^T h(\nu, t)\gamma_{id}(t - \nu, t)\, d\nu + \gamma_{dd}(t, t). \qquad (7) \end{aligned}$$

This expression gives the mean square ensemble system error for a particular time-variable linear system as specified by its weighting function $h(\tau, t)$. For fixed known nonstationary statistical correlation parameters, $\gamma_{ii}(t_1, t_2)$, $\gamma_{dd}(t_1, t_2)$, and $\gamma_{id}(t_1, t_2)$, it is required to minimize $\langle {}^je^2(t)\rangle_{\text{Av}}$ as a function of $h(\tau, t)$ in order to determine that time-variable linear system giving the least possible system error. (The existence and uniqueness of solution is tacitly assumed in above statement and others to follow. This is usually guaranteed from the physical nature of the problems under investigation.) The special $h(\tau, t)$ that gives the minimum value is called the optimum weighting function and characterizes the optimum system. If a particular $h(\tau, t)$ is optimum, then replacing $h(\tau, t)$ by a different operator $h(\tau, t) + \eta g(\tau, t)$, with arbitrary real η, cannot result in a smaller value. This variational technique proves the following.

Theorem 1

A necessary and sufficient condition that $h(\tau, t)$ be the optimum weighting function is that it satisfy the integral equation

$$\gamma_{id}(t - \nu, t) = \int_0^T h(\tau, t)\gamma_{ii}(t - \nu, t - \tau)\, d\tau, \qquad 0 \leq \nu \leq T. \qquad (8)$$

Proof: See Appendix.

Theorem 2

The minimum mean square ensemble system error resulting from the optimum choice of $h(\tau, t)$ is

$$\langle {}^je^2(t)\rangle_{\text{Av}} = \gamma_{dd}(t, t) - \int_0^T h(\tau, t)\gamma_{id}(t - \tau, t)\, d\tau. \qquad (9)$$

Proof: See Appendix.

Two major problems occur in practical applications of this basic theory. The first is to solve the appropriate integral equation (Theorem 1) using the particular correlation functions involved in the physical situation. The answer depends only on these statistical parameters and the time of operation, and consequently applies to all other applications having similar information. Secondly an engineering mechanization must be developed to reflect the mathematical result. A suggested design technique which will be valid for any desired operating region is to synthesize towards the required minimum ensemble system error (Theorem 2).

more extensive discussion of physical and mathatical ideas underlying this theory is contained in a vious paper.[6]

Section II. Integral Equation Solution for General Problem

he basic theory will now be applied to a general blem. For the input signals $\{{}^i s(t)\}$, any nonstationary dom process expressible in the following finite Fourier es ensemble form is allowed:

$$\begin{aligned}) &= \sum_{n=1}^{N} {}^i a_n \cos n\omega t + \sum_{n=1}^{N} {}^i b_n \sin n\omega t, \qquad t \geq 0 \\ &= 0, \qquad t < 0 \end{aligned} \tag{10}$$

ere ω = known constant, N = known integer, and

$$\begin{aligned} \langle {}^i a_n\, {}^i a_m \rangle_{\text{Av over } i} &= \alpha_{nm} \\ \langle {}^i b_n\, {}^i b_m \rangle_{\text{Av over } i} &= \beta_{nm} \\ \langle {}^i a_n\, {}^i b_m \rangle_{\text{Av over } i} &= \gamma_{nm} \end{aligned} \tag{11}$$

assumed to be known or calculable statistical quantities. e usual Fourier series constant term may be included, desired, by summing the series from 0 to N. Also, re is no restriction as to the relative number of sine or ine terms since the coefficients may be zero. The nals being zero for negative time many correspond ysically to the presence of a starting switch in the stem.

The desired outputs $\{{}^i d(t)\}$ are obtained from the inming signals by means of some preassigned time rying linear operation, so that for a suitable known ighting function $g(\tau, t)$, and for a time of operation m 0 to t,

$$\begin{aligned} {}^i d(t) &= \int_0^t g(\tau, t)\, {}^i s(t - \tau)\, d\tau, \qquad t \geq 0 \\ &= 0, \qquad t < 0. \end{aligned} \tag{12}$$

The noise random process $\{{}^i n(t)\}$ is assumed to have following autocorrelation function:

$$\begin{aligned} (t_1, t_2) &= A \exp(-k \mid t_1 - t_2 \mid) \cos c(t_1 - t_2), \\ &\qquad t_1, t_2 \geq 0 \\ &= 0, \qquad t_1 < 0 \quad \text{or} \quad t_2 < 0, \end{aligned} \tag{13}$$

k and c being non-negative known constants.

The various cross-correlations between noise and signal, noise and desired output, are assumed to be identically o.

The input message ${}^i i(t)$ is taken to be the sum of signal d noise,

$${}^i i(t) = {}^i s(t) + {}^i n(t). \tag{14}$$

Hence

$$\begin{aligned} \gamma_{ii}(t_1, t_2) &= \gamma_{ss}(t_1, t_2) + \gamma_{nn}(t_1, t_2) \\ &\qquad \text{when} \quad \gamma_{sn}(t_1, t_2) = 0 = \gamma_{ns}(t_1, t_2), \\ \gamma_{id}(t_1, t_2) &= \gamma_{sd}(t_1, t_2) \quad \text{when} \quad \gamma_{nd}(t_1, t_2) = 0. \end{aligned} \tag{15}$$

The general problem is now reduced to solving the integral (8) of Theorem 1, (with T replaced by t), for the optimum weighting function $h(\tau, t)$. Upon substituting the various correlation functions given above, the integral equation takes the form,

$$\begin{aligned} &\sum_1^N P(n, t) \cos n\omega(t - \nu) + \sum_1^N Q(n, t) \sin n\omega(t - \nu) \\ &\quad = \int_0^t h(\tau, t)\Big[\sum_1^N F(n, t - \tau) \cos n\omega(t - \nu) \\ &\qquad + \sum_1^N G(n, t - \tau) \sin n\omega(t - \nu) \Big]\, d\tau \\ &\qquad + \int_0^t h(\tau, t)[A e^{-k|\nu - \tau|} \cos c(\nu - \tau)]\, d\tau, \\ &\qquad\qquad 0 \leq \nu \leq t \end{aligned} \tag{16}$$

where everything is known except $h(\tau, t)$.

$$\begin{aligned} F(n, t - \tau) &= \sum_{m=1}^{N} [\alpha_{nm} \cos m\omega(t - \tau) \\ &\qquad + \tau_{nm} \sin m\omega(t - \tau)] \\ G(n, t - \tau) &= \sum_{m=1}^{N} [\beta_{nm} \sin m\omega(t - \tau) \\ &\qquad + \gamma_{mn} \cos m\omega(t - \tau)] \end{aligned} \tag{17}$$

$$\begin{aligned} P(n, t) &= \int_0^t g(\tau, t) F(n, t - \tau)\, d\tau \\ Q(n, t) &= \int_0^t g(\tau, t) G(n, t - \tau)\, d\tau. \end{aligned} \tag{18}$$

Outline of Solution

An exact mathematical solution of this complicated integral equation is obtained by setting up appropriate functions that will be amenable to Laplace transform treatment, and by creating systems of simultaneous time varying equations. In the Laplace transform analysis, all boundary conditions are ignored until the end when compensating delta functions are introduced. In order to solve for unknown time variable coefficients, sets of simultaneous time varying equations are derived which make the final result self-checking.

The main lines of the solution are as follows.

Working from the integral equation (16), let

$$\begin{aligned} I(\nu, t) &= A \int_0^t h(\tau, t) e^{-k|\nu - \tau|} \cos c(\nu - \tau)\, d\tau \\ &= \sum_1^N \mu_n(t) \cos n\omega(t - \nu) + \sum_1^N \nu_n(t) \sin n\omega(t - \nu) \end{aligned} \tag{19}$$

[6] J. S. Bendat, "A general theory of linear prediction and filter-," The Ramo-Wooldridge Corp., Control Systems Div.; Decem-, 1955. Also *J. Soc. Indust. Appl. Math.*, vol. 4, pp. 131–151; tember, 1956.

where

$$\mu_n(t) = P(n, t) - \int_0^t h(\tau, t)F(n, t - \tau)\, d\tau$$
$$\nu_n(t) = Q(n, t) - \int_0^t h(\tau, t)G(n, t - \tau)\, d\tau. \qquad (20)$$

The absolute value sign in the first expression for $I(\nu, t)$ is removed by breaking up into two integrals. An integration by parts is performed and variables changed so as to generate convolution product integrals. Eq. (19) is then solved for $h(\tau, t)$ in terms of $\mu_n(t)$ and $\dot{\nu}_n(t)$, ignoring the fact that $\mu_n(t)$ and $\nu_n(t)$ are themselves functions of $h(\tau, t)$! In functional form, after a number of such operations the result is

$$h(\tau, t) = \sum_1^N M_n(t) \cos n\omega\tau + \sum_1^N R_n(t) \sin n\omega\tau + S(t)e^{-a\tau}$$
$$+ Q(t)e^{a\tau} + U(t)\delta(\tau) + V(t)\delta(t - \tau),$$
$$a = (c^2 + k^2)^{1/2}, \qquad (21)$$

where $\delta(\tau)$ and $\delta(t - \tau)$ are delta functions and $M_n(t)$, $R_n(t)$, $S(t)$, $Q(t)$, $U(t)$, and $V(t)$ are $2N + 4$ unknown time varying coefficients still to be determined. These $2N + 4$ time varying coefficients are now found by substituting the functional form of $h(\tau, t)$, as shown in (21), into the expressions for $\mu_n(t)$ and $\nu_n(t)$, given by (20), and equating both sides of (19). This gives a system of $2N + 4$ simultaneous time varying equations which can be implicitly solved by following a special order. A more complete discussion of details appears in the Appendix.

Using the optimum weighting function (21), the minimum mean square ensemble system error, obtained from formula (9), is

$$\langle {}^ie^2(t)\rangle_{\text{Av}} = \sum_{n=1}^N P(n, t) \int_0^t [g(\tau, t) - h(\tau, t)] \cos n\omega(t - \tau)\, d\tau$$
$$+ \sum_{n=1}^N Q(n, t) \int_0^t [g(\tau, t) - h(\tau, t)] \sin n\omega(t - \tau)\, d\tau. \qquad (22)$$

Section III. Two Applications

The incoming signal, considered to be a member of a random process $\{{}^is(t)\}$, is of the form

$$\begin{aligned} {}^is(t) &= {}^ia \cos \omega t + {}^ib \sin \omega t, & t \geq 0 \\ &= 0, & t < 0 \end{aligned} \qquad (23)$$

where $\omega =$ constant,

$$\begin{aligned} \langle {}^ia^2\rangle_{\text{Av over } i} &= \langle {}^ib^2\rangle_{\text{Av over } i} = \alpha \\ \langle {}^ia\, {}^ib\rangle_{\text{Av over } i} &= 0 \end{aligned} \qquad (24)$$

so that the autocorrelation function $\gamma_{ss}(t_1, t_2)$ becomes

$$\begin{aligned} \gamma_{ss}(t_1, t_2) &= \alpha \cos \omega(t_1 - t_2), & t_1, t_2 \geq 0 \\ &= 0, & t_1 < 0 \text{ or } t_2 < 0. \end{aligned} \qquad (25)$$

This signal is a special case (where $N = 1$) of the more general signals treated in the previous section.

The noise random process $\{{}^in(t)\}$ is assumed to have the same damped exponential-cosine autocorrelation function as before, (13). It is also again supposed that the various cross-correlation functions are zero.

There are two objectives in mind. In Case 1, the desired output is to recover the input signal, and in Case 2, it is required to approximate the integral of the incoming signal. Thus, to be specific, if the incoming message is a distorted velocity signal, Case 1 has the task of finding the optimum weighting function for a filter recovering the velocity while Case 2 determines the optimum filter to know the position.

Case 1

$${}^jd(t) = {}^js(t) \quad \text{for all} \quad j. \qquad (26)$$

Therefore,

$$\gamma_{dd}(t_1, t_2) = \gamma_{sd}(t_1, t_2) = \gamma_{ss}(t_1, t_2). \qquad (27)$$

The integral equation which needs to be solved for the optimum weighting function $h(\tau, t)$ from (16), is

$$\alpha \cos \omega\nu = \int_0^t h(\tau, t)[\alpha \cos \omega(\nu - \tau)$$
$$+ Ae^{-k|\nu-\tau|} \cos c(\nu - \tau)]\, d\tau \qquad 0 \leq \nu \leq t. \qquad (28)$$

Omitting all details, the solution of (28) may be found from the previous analysis (with $N = 1$ and statistical terms as indicated above). The solution is

$$h(\tau, t) = M(t) \cos \omega\tau + R(t) \sin \omega\tau + S(t)e^{-a\tau}$$
$$+ Q(t)e^{a\tau} + U(t)\delta(\tau) + V(t)\delta(t - \tau),$$
$$a = (c^2 + k^2)^{1/2}, \qquad (29)$$

where the time varying coefficients $M(t)$, $R(t)$, $S(t)$, $U(t)$, and $V(t)$ are determined through a known system of six simultaneous time varying equations. On solving these simultaneous equations, one can demonstrate that for large t,

$$\begin{aligned} M(t) &\sim 2(t + \lambda)^{-1} & &\text{as} \quad t \to \infty \\ R(t) &= o(M(t)) & &\text{as} \quad t \to \infty \\ S(t) &\sim \rho M(t) & &\text{as} \quad t \to \infty \\ Q(t) &= o(e^{-at}) & &\text{as} \quad t \to \infty \\ U(t) &\sim \rho_1 M(t) & &\text{as} \quad t \to \infty \\ V(t) &\sim (\rho_2 \sin \omega t + \rho_3 \cos \omega t)M(t) & &\text{as} \quad t \to \infty \end{aligned} \qquad (30)$$

where the notation $f(t) \sim g(t)$ read "$f(t)$ asymptotic to $g(t)$", as $t \to \infty$, means that $f(t)/g(t) \to 1$ as $t \to \infty$, while $f(t) = o(g(t))$ as $t \to \infty$ means that $f(t)/g(t) \to 0$ as $t \to \infty$. The parameters λ, ρ, ρ_1, ρ_2 and ρ_3 denote certain constants t expressions for $M(t)$ and $R(t)$ for all times t may be ined if desired, and through them exact expressions he other time varying coefficients. Unfortunately, are lengthy expressions involving combinations of nometric and exponential functions which are much cumbersome to be useful. The precise nature of $_2$ and ρ_3 is not significant. However, λ and ρ have the wing meaning in terms of known information,

$$\lambda = 4Ak(\omega^2 + a^2)(\alpha d)^{-1} \tag{31}$$

$$\rho = 2a(a - k)(\omega^2 - a^2)d^{-1} \tag{32}$$

$$d = [(\omega - c)^2 + k^2][(\omega + c)^2 + k^2] \tag{33}$$

$$a = (c^2 + k^2)^{1/2}. \tag{34}$$

that $\lambda > 0$ if $Ak > 0$, and $\rho = 0$ if $a - k = 0$. The tant $a - k = 0$ if $c = 0$, namely, when the noise correlation function is a damped exponential alone. rom (22), the minimum mean square ensemble em error for arbitrary values of t, when using the num weighting function (29) is given by

$$\langle {}^i e^2(t)\rangle_{\text{Av}} = 2Ak(\omega^2 + a^2)d^{-1}M(t). \tag{35}$$

error approaches zero for large t since $M(t)$ has this erty.

2

all j,

$$= \int_0^t {}^i s(\tau)\, d\tau$$

$$= \begin{cases} {}^i a\omega^{-1} \sin \omega t + {}^i b\omega^{-1}(1 - \cos \omega t), & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{36}$$

new integral equation which one needs to solve for the mum weighting function $h(\tau, t)$ is

$$[\sin \omega\nu + \sin \omega(t - \nu)] = \alpha \int_0^t h(\tau, t) \cos \omega(\nu - \tau)\, d\tau$$

$$+ A \int_0^t h(\tau, t) e^{-k|\nu - \tau|} \cos c(\nu - \tau)\, d\tau,$$

$$0 \leq \nu \leq t. \tag{37}$$

efore, the solution has the form

$$t) = M(t) \cos \omega\tau + R(t) \sin \omega\tau + S(t)e^{-a\tau}$$

$$+ Q(t)e^{a\tau} + U(t)\delta(\tau) + V(t)\delta(t - \tau),$$

$$a = (c^2 + k^2)^{1/2} \tag{38}$$

re the time varying coefficients $S(t)$, $Q(t)$, $U(t)$ and bear similar relationships to $M(t)$ and $R(t)$ as in e 1. It is mainly in the final determination of $M(t)$, and $S(t)$ that Cases 1 and 2 differ, for now,

$$\begin{aligned} I(t) &\sim 2[\omega(t + \lambda)]^{-1} \sin \omega t && \text{as} \quad t \to \infty \\ R(t) &\sim 2[\omega(t + \lambda)]^{-1}[1 - \cos \omega t] && \text{as} \quad t \to \infty \\ S(t) &\sim \rho M(t) + \sigma R(t) && \text{as} \quad t \to \infty. \end{aligned} \tag{39}$$

The parameters λ and ρ are the same as previously, (31) and (32), while the constant σ is given by

$$\sigma = 4\omega ak(a - k)d^{-1}. \tag{40}$$

Note that $\sigma = 0$ if $a - k = 0$, that is, if $c = 0$. If desired, exact lengthy expressions may be derived for the time varying coefficients at arbitrary operating times t.

The minimum mean seuare ensemble system error possible in Case 2, for any value of t, is given by

$$\langle {}^i e^2(t)\rangle_{\text{Av}} = 2Ak(\omega^2 + a^2((\omega d)^{-1}$$

$$\cdot [M(t) \sin \omega t + R(t)(1 - \cos \omega t)] \tag{41}$$

As in Case 1, this error approaches zero for large t.

Section IV. Synthesis of Separable Time Variable Weighting Functions

From the nature of the solutions derived in the first part of the paper, (21), a new category of separable time variable weighting functions will be defined and investigated.

Definition 1

A time variable weighting function $h(\tau, t)$ is said to be separable if

$$h(\tau, t) = f(t)g(\tau) \tag{42}$$

where $f(t)$, the time varying factor, is a function of t alone, while $g(\tau)$, the constant parameter factor, is a function of τ alone.

The decomposition possible for separable time variable weighting functions is essential to the further discussion. In particular, more complicated separable weighting functions may be involved of the form,

$$h(\tau, t) = \sum_{n=1}^{M} f_n(t)g_n(\tau) + U(t)\delta(\tau) + V(t)\delta(t - \tau) \tag{43}$$

where M = integer, $\delta(\tau)$ and $\delta(t - \tau)$ are delta functions at $\tau = 0$ and $\tau = t$, respectively.

The response $r(t)$ to an input $i(t)$ for a system characterized by a separable weighting function (43) is given by

$$r(t) = \sum_{n=1}^{M} f_n(t) \int_0^t g_n(\tau)i(t - \tau)\, d\tau + U(t)i(t) + V(t)i(0). \tag{44}$$

This follows directly from the definition of $h(\tau, t)$ and the superposition property of linear systems. Note that the effect of the delta function terms is to merely pick out the final and initial values of the input for special attention. These are multiplied by time varying gains $U(t)$ and $V(t)$. The factors $g_n(\tau)$, $n = 1, 2, \cdots, M$ represent M constant parameter linear weighting functions, and the corresponding M time varying coefficients $f_n(t)$, $n = 1, 2, \cdots, M$ denote time varying gain amplifiers.

A schematic diagram to generate $h(\tau, t)$ for arbitrary operating times is displayed in Fig. 1. Switch S is initially closed at $t = 0$ and open for all $t > 0$.

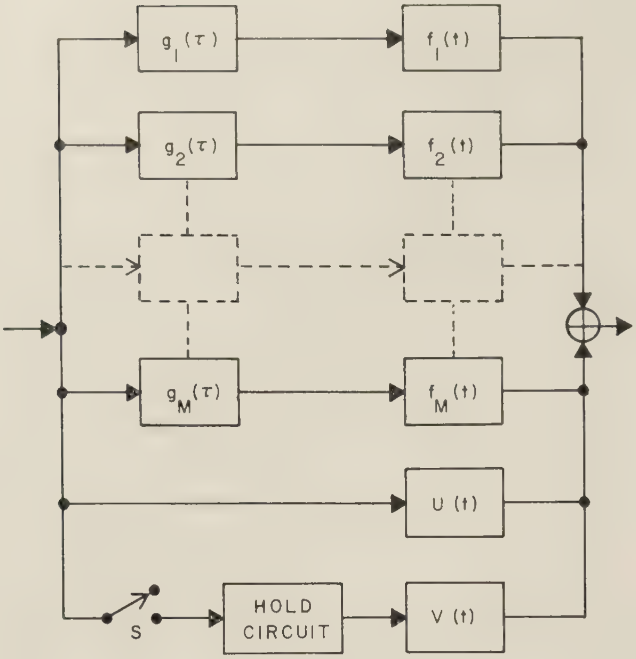


Fig. 1—Exact synthesis of separable $h(\tau, t)$.

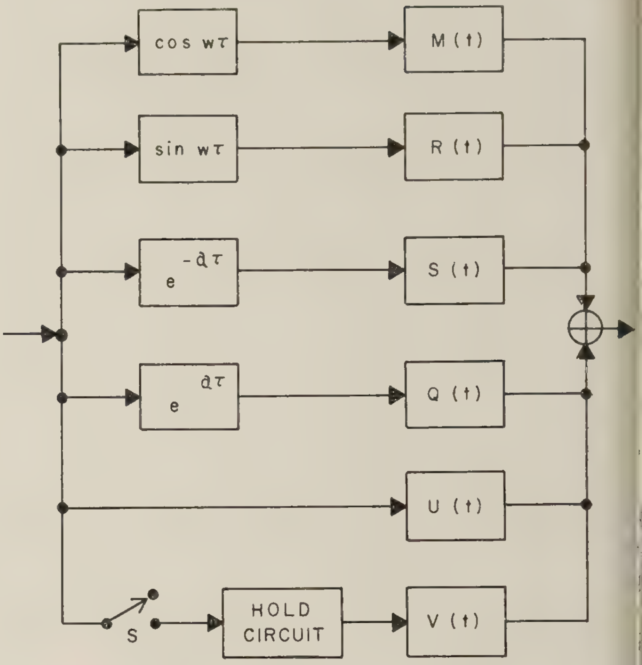


Fig. 2—Exact synthesis for $N = 1$ examples.

Fig. 1 indicates how to synthesize any separable time variable weighting function of class (43). If one now examines important features of the optimum time variable weighting functions determined in (21) for a general case, $M = 2N + 2$ and the constant parameter factors $g_n(\tau)$, $n = 1, 2, \cdots, 2N + 2$, are of the following definite types:

$$g_n(\tau) = \begin{cases} \cos n\omega\tau, & \sin n\omega\tau, \quad n = 1, 2, \cdots, N \\ e^{-a\tau}, & e^{a\tau}, \quad a = (c^2 + k^2)^{1/2}. \end{cases} \tag{45}$$

The optimum time variable linear weighting function is given by

$$h_0(\tau, t) = \sum_1^N M_n(t) \cos n\omega\tau + \sum_1^N R_n(t) \sin n\omega\tau + S(t)e^{-a\tau} + Q(t)e^{a\tau} + U(t)\delta(\tau) + V(t)\delta(t - \tau) \tag{46}$$

where the $2N + 4$ time varying coefficients $M_n(t)$, $R_n(t)$, $S(t)$, $Q(t)$, $U(t)$ and $V(t)$ are found from a system of $2N + 4$ simultaneous time varying equations (details in Appendix). Simple mechanizations for the constant parameter weighting functions of (45) are well known.

In the two examples of Section III, the optimum time variable weighting function has the form corresponding to $N = 1$,

$$h_0(\tau, t) = M(t) \cos \omega\tau + R(t) \sin \omega\tau + S(t)e^{-a\tau} + Q(t)e^{a\tau} + U(t)\delta(\tau) + V(t)\delta(t - \tau). \tag{47}$$

The six time varying coefficients $M(t)$, $R(t)$, $S(t)$, $Q(t)$, $U(t)$, and $V(t)$ are determined from a system of six simultaneous time varying equations. Consequently, if these six quantities are mechanizable, an exact synthesis for arbitrary times of operation would be given by the schematic design of Fig. 2.

Figs. 1 and 2 show useful direct ways for handlir many synthesis problems. Fig. 1 illustrates how o mechanizes a general separable time variable weightir function (43), while Fig. 2 gives a design appropriate t the optimum time variable linear filters derived in Sectic III for the special cases when $N = 1$. Optimum filters f arbitrary $N \geq 2$ are obtained by inserting additional c $n\omega\tau$ and sin $n\omega\tau$ components, $n = 2, 3, \cdots, N$, multiplie by required time varying gains $M_n(t)$ and $R_n(t)$.

Definition 2

For long-time (short-time) operation, $h_a(\tau, t)$ is a asymptotic weighting function to an optimum weightir function $h_0(\tau, t)$ if

$$h_a(\tau, t)/h_0(\tau, t) \rightarrow 1 \quad \text{as} \quad t \rightarrow \infty \ (t \rightarrow 0). \qquad (4$$

This is written

$$h_a(\tau, t) \sim h_0(\tau, t) \quad \text{as} \quad t \rightarrow \infty \ (t \rightarrow 0).$$

An immediate consequence of this definition is that f separable $h_0(\tau, t)$, say, $h_0(\tau, t) = f_0(t)\, g_0(\tau)$, then $h_a(\tau, t)$ $f_a(t)\, g_0(\tau) \sim h_0(\tau, t)$ as $t \rightarrow \infty$ $(t \rightarrow 0)$ according to wheth or not $f_a(t) \sim f_0(t)$ as $t \rightarrow \infty$ $(t \rightarrow 0)$. Also, for the extende class of separable time variable weighting functio covered by (43), asymptotic forms can be derived b merely finding suitable asymptotic forms to each of t time varying coefficients. This is a particularly convenie method to follow for optimum weighting functions (4 where the time varying coefficients are specified throu systems of simultaneous equations.

Definition 3

The notation $r_0(t)$ and $r_a(t)$ will be used to denote t response of an optimum system $h_0(\tau, t)$ and its asymptot tem $h_a(\tau, t)$, respectively, to an arbitrary input $i(t)$.

$$r_0(t) = \int_0^t h_0(\tau, t) i(t - \tau)\, d\tau \tag{49}$$

$$r_a(t) = \int_0^t h_a(\tau, t) i(t - \tau)\, d\tau. \tag{50}$$

The fact that $h_a(\tau, t)/h_0(\tau, t) \to 1$ necessarily only as ∞ $(t \to 0)$, plus the observation that $| r_0(t) |$ is bounded a bounded input proves $| r_a(t) - r_0(t) | \to 0$ as $t \to \infty$ $\to 0)$. This leads to

eorem 3

Mechanization of an asymptotic weighting function es a synthesis which is valid in the corresponding mptotic region of operation. Nothing can be inferred general about its performance during the non-mptotic period. Proof: See Appendix.

A synthesis for an asymptotic separable weighting ction, satisfying (48), will have a schematic design ilar to Fig. 1 or Fig. 2.

CTION V. EXAMPLES OF ASYMPTOTIC SYSTEM DESIGN

Valid mechanizations for long-time operation will now derived for the optimum time variable linear filter ghting functions found in Cases 1 and 2. Exact synthesis all times of operation does not appear to be a feasible gineering problem.

se 1

As $t \to \infty$, one sees from (29) and (30) that the asymp-ic weighting function takes the form

$$\tau, t) = 2(t + \lambda)^{-1}[\cos \omega\tau + \rho e^{-a\tau} + \rho_1 \delta(\tau) + (\rho_2 \sin \omega t + \rho_3 \cos \omega t)\delta(t - \tau)]. \tag{51}$$

ause of the time factor $(t + \lambda)^{-1}$, the effects of the delta functions (*i.e.*, merely picking out the initial final values of the input) can be ignored in long-time ration. Hence, one would synthesize only the nondelta ction terms,

$$h_a(\tau, t) = 2(t + \lambda)^{-1}[\cos \omega\tau + \rho e^{-a\tau}]. \tag{52}$$

appropriate design for the asymptotic weighting ction (52) is drawn in Fig. 3.

The system error $e(t)$ associated with use of the above cuit will now be calculated. From formula (2), together h (14), and (52), omitting the ensemble superscript j,

$$= 2(t + \lambda)^{-1} \int_0^t [\cos \omega\tau + \rho e^{-a\tau}] \cdot [s(t - \tau) + n(t - \tau)]\, d\tau - s(t). \tag{53}$$

is is integrated easily and shows for large t that $\to e_a(t)$ where

$$) = 2(t + \lambda)^{-1} \int_0^t [\cos \omega\tau + \rho e^{-a\tau}] n(t - \tau)\, d\tau. \tag{54}$$

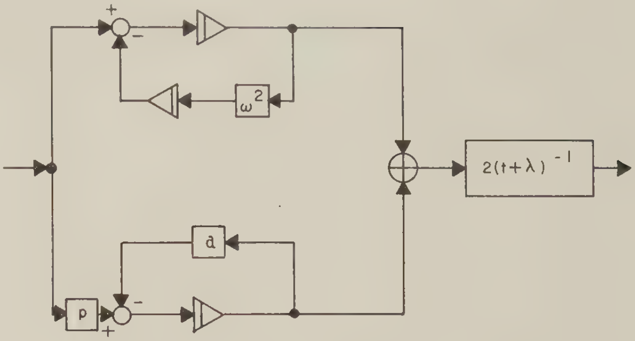


Fig. 3—Case 1 filter for long-time operation.

The quantity $e_a(t)$ represents the system error during the desired asymptotic long-time region of operation. Note the two properties:

1) $e_a(t)$ is a function of $n(t)$ alone such that $e_a(t) \to 0$ if $n(t) \to 0$.
2) $e_a(t)$ and its first derivative $\dot{e}_a(t)$ are bounded in absolute value and, in fact, decrease with time under the hypothesis that $n(t)$ is bounded.

Case 2

In Case 2, from (38) and (39), the optimum time variable linear filter weighting function for large t has an asymptotic form (exclusive of impulse terms) given by

$$h_a(\tau, t) = 2[\omega(t + \lambda)]^{-1}\{\sin \omega t[\cos \omega\tau + \rho e^{-a\tau}] + (1 - \cos \omega t)[\sin \omega\tau + \sigma e^{-a\tau}]\}. \tag{55}$$

Again, the impulse functions give a negligible contribution for long-time operation, and are therefore ignored. A schematic design for this asymptotic weighting function is shown in Fig. 4.

From (2), (36), and (55), the system error $e(t)$ corresponding to the Case 2 circuit is determined. For large t, this system error $e(t) \to e_a(t)$ where

$$e_a(t) = \int_0^t h_a(\tau, t) n(t - \tau)\, d\tau. \tag{56}$$

Thus, the same two properties of $\dot{e}_a(t)$ relative to $n(t)$ are seen to hold for the Case 2 circuit as the Case 1 circuit. In addition, because of the factors $\sin \omega t$ and $(1 - \cos \omega t)$ in $h_a(\tau, t)$ of Case 2 [see. (55)], it is clear that the asymptotic system error $e_a(t)$ [see (56)] will be zero at regular periodic intervals of time $t = 2n\pi\omega^{-1}$, $(n = 0, 1, 2, 3, \cdots)$.

CONCLUSION

A few words about the fundamental nature of this work. First of all, the one serious restriction on the incoming signals is a possible lack of available statistical information. Improved computing machinery and data handling techniques are helping to fill this need. The class of random processes expressible in finite Fourier series ensemble form, to which the incoming signals and desired outputs belong, covers many technical problems. For example, the input signals form a Gaussian random process

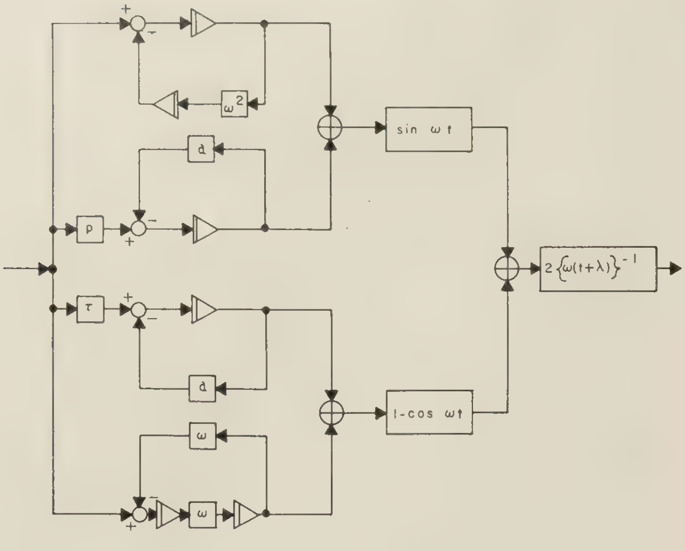


Fig. 4—Case 2 filter for long-time operation.

if the statistics are such that $\langle {}^i a_n^2 \rangle_{\text{Av over } i} = \langle {}^i b_n^2 \rangle_{\text{Av over } i}$, independent of n, while all other double moments are zero, and if the distribution of the random variables $\{{}^i a_n\}$ and $\{{}^i b_n\}$ is a normal distribution for all n. This case has been frequently studied. Secondly, it is an empirical fact that the noise perturbation treated, one whose autocorrelation function combines periodicity with exponential decay, occurs in many diverse places. Frequently, the autocorrelation function may be fitted to a damped exponential term alone. Thirdly, it should be emphasized that exact expressions for finite operating periods from 0 to any time t have been derived. Long-time operating period situations are, however, also included by letting t approach infinity. Thus, a solution has been given not for one special problem but for many important physical applications.

The two-step technique which was employed to solve the general integral equation appearing in this paper may be appropriate for other similar integral equations. In brief, this procedure requires that one should determine the functional form of the optimum weighting function except for unknown time varying coefficients, and then substitute this functional form back into the original integral equation so as to yield sets of simultaneous time varying equations whose solution gives the desired coefficients. A derivation so obtained is automatically self-checking. It is not required that the first step be carried out through use of Laplace transform theory, as was done in this work. Instead, one is free to use whatever insight he might have in a particular situation to propose a possible functional form, and proceed with the second step of satisfying the integral equation. Additional terms would be added, or substracted, depending on the derived results.

In particular, this two-step technique should be used whenever the noise autocorrelation function is a linear combination of exponential or exponential-cosine autocorrelation functions, instead of a single such term as is considered in the paper. Since it is well known that many empirical autocorrelation functions may be fitted to such a linear combination of basic autocorrelation functions without difficulty, the analysis is thus able to cover these more general situations.

The class of separable time variable weighting functions, whose concept is abstracted from the form of the integral equation solutions, is deemed by the author to be an important physical entity. The synthesis techniques discussed herein refer to this extended class of separable filters and are not limited to the optimum filters alone. Some open problems remain in determining under what restrictive conditions the optimal weighting function is of this separable form. Here, it has been demonstrated to occur for a certain large class of optimization problems, but the question of its presence elsewhere is not known.

The asymptotic synthesis ideas offer a fresh insight to the mechanization problem when exact synthesis is difficult and when asymptotic operating regions are involved. A resulting design is necessarily valid only in the corresponding asymptotic region of operation. Consequently, if the apparatus is also to be used in non-asymptotic regions, extensive tests and engineering adjustments would be required before final production. Some tests, of course, should always be made to compare performance of a theoretically proposed system with, possibly, simpler circuits motivated by engineering experience.

Appendix

Proof of Theorems 1 and 2

Let

$$E^2(t) = \langle {}^i e^2(t) \rangle_{\text{Av}}$$
$$= \int_0^T \int_0^T h(\nu, t) h(\tau, t) \gamma_{ii}(t-\nu, t-\tau)\, d\tau\, d\nu \qquad (57)$$
$$- 2 \int_0^T h(\nu, t) \gamma_{id}(t-\nu, t)\, d\nu + \gamma_{dd}(t, t).$$

Assume $h(\tau, t)$ is the optimum weighting function. Let $N^2(t)$ be the mean square ensemble error using another operator $h(\tau, t) + \eta g(\tau, t)$ where η is an arbitrary real constant.

$$N^2(t) = \int_0^T \int_0^T [h(\nu, t) + \eta g(\nu, t)][h(\tau, t) + \eta g(\tau, t)]$$
$$\cdot \gamma_{ii}(t-\nu, t-\tau)\, d\tau\, d\nu \qquad (58)$$
$$- 2 \int_0^T [h(\nu, t) + \eta g(\nu, t)]$$
$$\cdot \gamma_{id}(t-\nu, t)\, d\nu + \gamma_{dd}(t, t)$$
$$= E^2(t) - 2\eta M(t) + \eta^2 L(t) \qquad (59)$$

where

$$M(t) = \int_0^T g(\nu, t) \Big[\gamma_{id}(t-\nu, t)$$
$$- \int_0^T h(\tau, t) \gamma_{ii}(t-\nu, t-\tau)\, d\tau \Big]\, d\nu \qquad (60)$$

$$L(t) = \int_0^T g(\nu, t)g(\tau, t)\gamma_{ii}(t - \nu, t - \tau)\, d\tau\, d\nu$$
$$= \left\langle \left[\int_0^T g(\tau, t)\, {}^i i(t - \tau)\, d\tau \right]^2 \right\rangle_{Av} \geq 0. \tag{61}$$

he result of Theorem 1 follows from showing that) is a minimum if and only if $M(t)$ [see (60)] is identi- zero for any choice of the operator $g(\nu, t)$, $0 \leq \nu \leq T$. $M(t) \equiv 0$ for any $g(\nu, t)$, $0 \leq \nu \leq T$, then $N^2(t) =$) $+ \eta^2 L(t) \geq E^2(t)$ for any η, since $L(t) \geq 0$. There- , $E^2(t)$ is a minimum. This proves the sufficiency of orem 1.

onversely, if $M(t) \not\equiv 0$ for all $g(\nu, t)$, $0 \leq \nu \leq T$, then can make $M(t) > 0$ by changing the sign of a particular $t)$ if required. Now,

$$N^2(t) = E^2(t) - 2\eta[M(t) - (\eta/2)L(t)] \tag{62}$$

η sufficiently small and positive, $M(t) - (\eta/2)$ $t) > 0$. Then $N^2(t) < E^2(t)$, giving a contradiction to original claim that $E^2(t)$ was a minimum value. This ves the necessity of Theorem 1.

he statement of Theorem 2 results by substituting integral equation requirement of Theorem 1 into original expression (57) for $E^2(t)$.

ctional Form of $h(\tau, t)$

tarting from (19) and (20), one can demonstrate after eral steps that

$$- x, t) = Ae^{-kx} \int_0^t h(y, t)e^{-k(t-y)} \cos c(t - x - y)\, dy$$
$$+ 2Ak \int_0^x e^{-k(x-\phi)} \left[\int_0^{t-\phi} h(y, t) \cdot e^{-k(t-\phi-y)} \cos c(t - x - y)\, dy \right] d\phi \tag{63}$$
$$= \sum_1^N \mu_n(t) \cos n\omega x + \sum_1^N \nu_n(t) \sin n\omega x. \tag{64}$$

ine

$$_1(t - \nu, t) = \int_0^\nu h(y, t)e^{-k(\nu-y)} \cos c(\nu - y)\, dy$$
$$_2(t - \nu, t) = \int_0^\nu h(y, t)e^{-k(\nu-y)} \sin c(\nu - y)\, dy. \tag{65}$$

et $\lambda(s, t)$ be the Laplace transform of $\lambda(x, t)$ with pect to x, t being held constant; similarly, let $- \rho, t)$ be the Laplace transform of $\lambda(t - \nu, t)$ with pect to ν, t being held constant. Upon taking the place transform of (63) and (64) with respect to x, er substituting (65), and using the relation $A\lambda_1(0, t) =$ $\mu_n(t)$ which results from (63) when $x = 0$, one obtains

$$k[(s + k)\lambda_1(s, t) + c\lambda_2(s, t)]$$
$$= L(t) + \sum_1^N [sF_n(t) + n\omega G_n(t)][s^2 + n^2\omega^2]^{-1} \tag{66}$$

where the forms of $L(t)$, $F_n(t)$ and $G_n(t)$ are not important. The Laplace transforms of (65) with respect to ν, t being held constant, are

$$\lambda_1(t - p, t) = H(p, t)[p + k][(p + k)^2 + c^2]^{-1}$$
$$\lambda_2(t - p, t) = H(p, t)[c][(p + k)^2 + c^2]^{-1} \tag{67}$$

where $H(p, t)$ represents the Laplace transform of the unknown weighting function $h(\nu, t)$ with respect to ν, t being held constant.

In order to solve for $H(p, t)$ from (67), it is necessary to change the Laplace transforms $\lambda(s, t)$ of (66) into expressions of type $\lambda(t - p, t)$. Take the inverse Laplace transform of (66) with respect to s, ignoring all boundary conditions. Then replace x in $\lambda(x, t)$ by $t - \nu$ and take new Laplace transforms with respect to ν. This yields the desired relation, and gives for $H(p, t)$,

$$H(p, t) = \sum_1^N [pM_n(t) + n\omega R_n(t)][p^2 + n^2\omega^2]^{-1}$$
$$+ (p + a)^{-1}S(t) + (p - a)^{-1}Q(t) \tag{68}$$

where

$$a = (c^2 + k^2)^{1/2}, \tag{69}$$

and $M_n(t)$, $R_n(t)$, $S(t)$ and $Q(t)$ are unknown time varying coefficients.

Finally, the inverse Laplace transform of (68), together with additional impulse factors $U(t)\,\delta(t)$ and $V(t)\,\delta(t - \tau)$ (which are introduced to account for previously neglected boundary conditions) shows that the optimum time variable linear weighting function $h(\tau, t)$ has the functional form,

$$h(\tau, t) = \sum_1^N M_n(t) \cos n\omega\tau + \sum_1^N R_n(t) \sin n\omega\tau + S(t)e^{-a\tau}$$
$$+ Q(t)e^{a\tau} + U(t)\delta(\tau) + V(t)\delta(t - \tau),$$
$$a = (c^2 + k^2)^{1/2}, \tag{70}$$

where, for ease in later computation, $\delta(\tau)$ will be defined as the usual (Dirac) delta function multiplied by two, that is

$$\delta(\tau) = 0 \qquad \text{for any} \quad \tau \neq 0$$
$$\int_0^t \delta(\tau)\, d\tau = \int_{-t}^0 \delta(\tau)\, d\tau = 1 \quad \text{for any} \quad t \neq 0. \tag{71}$$

Altogether, there are $2N + 4$ time varying coefficients which must be determined in (70).

Determination of TimeVarying Coefficients

For $h(\tau, t)$ as given by (70), let

$$J(\nu, t) = \int_0^t h(\tau, t)e^{-k|\nu-\tau|} \cos c(\nu - \tau)\, d\tau. \tag{72}$$

Then, clearly, $h(\tau, t)$ will be the optimum weighting function if

$$I(\nu, t) = AJ(\nu, t) \tag{73}$$

since the optimum weighting function must satisfy (19) and (20).

By equating corresponding terms on both sides of (73), the $2N + 4$ unknown time varying coefficients of (70) may be shown to satisfy the following system of $2N + 4$ simultaneous equations, whose consistency is guaranteed on the basis of physical validity of the processes under investigation.

$$\begin{aligned} 2Ak\Delta_n M_n(t) &= d_n[\mu_n(t)\cos n\omega t + \nu_n(t)\sin n\omega t] \\ &\qquad (n = 1, 2, \cdots, N) \\ 2Ak\Delta_n R_n(t) &= d_n[\mu_n(t)\sin n\omega t - \nu_n(t)\cos n\omega t) \\ &\qquad (n = 1, 2, \cdots, N) \end{aligned} \tag{74}$$

$$\begin{aligned} U(t) &= -\left(\sum_1^N b_n^{(1)} M_n(t) + \sum_1^N b_n^{(5)} R_n(t) \right. \\ &\qquad \left. + b_1 S(t) + b_5 Q(t)\right) \\ V(t) &= -\left(\sum_1^N b_n^{(3)} M_n(t) + \sum_1^N b_n^{(7)} R_n(t) \right. \\ &\qquad \left. + b_3 e^{-at} S(t) + b_7 e^{at} Q(t)\right) \end{aligned} \tag{75}$$

$$\begin{aligned} \sum_1^N b_n^{(2)} M_n(t) + \sum_1^N b_n^{(6)} R_n(t) + b_2 S(t) + b_6 Q(t) &= 0 \\ \sum_1^N b_n^{(4)} M_n(t) + \sum_1^N b_n^{(8)} R_n(t) + b_4 e^{-at} S(t) + b_8 e^{at} Q(t) &= 0 \end{aligned} \tag{76}$$

where

$$\begin{aligned} \Delta_n &= n^2\omega^2 + a^2, \qquad a^2 = c^2 + k^2, \\ d_n &= \Delta_n^2 - 4n^2\omega^2 c^2 \end{aligned} \tag{77}$$

$$\begin{aligned} d_n b_n^{(1)} &= -k\Delta_n, \qquad d_n b_n^{(3)} = n\omega(\Delta_n - 2c^2)\sin n\omega t \\ &\qquad\qquad - k\Delta_n \cos n\omega t \\ d_n b_n^{(2)} &= c(\Delta_n - 2n^2\omega^2), \qquad d_n b_n^{(4)} = 2n\omega ck \sin n\omega t \\ &\qquad\qquad + c(2n^2\omega^2 - \Delta_n)\cos n\omega t \end{aligned} \tag{78}$$

$$\begin{aligned} d_n b_n^{(5)} &= n\omega(\Delta_n - 2c^2), \qquad d_n b_n^{(7)} = -k\Delta_n \sin n\omega t \\ &\qquad\qquad + n\omega(2c^2 - \Delta_n)\cos n\omega t \\ d_n b_n^{(6)} &= -2n\omega ck, \qquad d_n b_n^{(8)} = c(2n^2\omega^2 - \Delta_n)\sin n\omega t \\ &\qquad\qquad - 2n\omega ck \cos n\omega t \end{aligned} \tag{79}$$

while

$$\begin{aligned} b_1 &= (2a)^{-1}, \qquad b_3 = -(2a)^{-1} \\ b_2 &= c[2a(a-k)]^{-1}, \qquad b_4 = -c[2a(a+k)]^{-1} \end{aligned} \tag{80}$$

$$\begin{aligned} b_5 &= -(2a)^{-1}, \qquad b_7 = (2a)^{-1} \\ b_6 &= c[2a(a+k)]^{-1}, \qquad b_8 = -c[2a(a-k)]^{-1}. \end{aligned} \tag{81}$$

Assuming $M_n(t)$ and $R_n(t)$ remain finite as t approaches infinity, it follows that $S(t)$ is finite as t approaches infinity while $Q(t)$ approaches zero. The $2N + 4$ time varying coefficients are determined by the following procedure: First, $S(t)$ and $Q(t)$ are found in terms of $M_n(t)$ and $R_n(t)$ by (76); then, $U(t)$ and $V(t)$ are known as functions of $M_n(t)$ and $R_n(t)$ by (75). Finally, $M_n(t)$ and $R_n(t)$ are calculated from the $2N$ simultaneous (74).

Proof of Theorem 3

Suppose $h_0(\tau, t)$ is of a separable form, for example,

$$h_0(\tau, t) = f_0(t)g_0(\tau) \quad \text{with} \quad h_a(\tau, t) = f_a(t)g_0(\tau) \sim h_0(\tau, t) \quad \text{as} \quad t \to \infty \ (t \to 0). \tag{82}$$

Then, $f_a(t) \sim f_0(t)$ as $t \to \infty$ $(t \to 0)$.

In practical cases of interest, a bounded input will produce a bounded output. Hence, there exists some constant $K \geq 0$ such that

$$| r_0(t) | = \left| f_0(t) \int_0^t g_0(\tau) i(t - \tau)\, d\tau \right| < K \quad \text{for all} \quad t \geq 0. \tag{83}$$

The difference,

$$| r_a(t) - r_0(t) | \leq | \{f_a(t)/f_0(t)\} - 1 | K \to 0 \quad \text{as} \quad t \to \infty \ (t \to 0). \tag{84}$$

This convergence need not take place outside the asymptotic region. Similarly, if

$$h_0(\tau, t) = \sum_{n=1}^{M} f_{n0}(t) g_n(\tau) + U_0(t)\delta(\tau) + V_0(t)\delta(t - \tau) \tag{85}$$

and

$$h_a(\tau, t) = \sum_{n=1}^{M} f_{na}(t) g_n(\tau) + U_a(t)\delta(\tau) + V_a(t)\delta(t - \tau) \tag{86}$$

where

$$\begin{aligned} f_{na}(t) &\sim f_{n0}(t) \quad \text{as} \quad t \to \infty \ (t \to 0), \quad n = 1, 2, \cdots, M \\ U_a(t) &\sim U_0(t) \quad \text{as} \quad t \to \infty \ (t \to 0), \\ V_a(t) &\sim V_0(t) \quad \text{as} \quad t \to \infty \ (t \to 0), \end{aligned} \tag{87}$$

then $r_a(t) \to r_0(t)$ necessarily only as $t \to \infty$ $(t \to 0)$.

Acknowledgment

This material was first studied at Northrop Aircraft, Inc., Hawthorne, Calif. In its initial formulation, it received limited circulation in a company report.[7] The author is grateful to his former colleagues at Northrop for their helpful criticisms. The present paper represents both a revision and extension of the earlier effort. The author wishes to thank the Ramo-Wooldridge Corporation, Los Angeles, Calif., for providing a stimulating atmosphere and facilities with which to complete the work.

[7] J. S. Bendat, "Optimum Time-Variable Filtering for Nonstationary Random Processes," Northrop Aircraft, Inc., Rep. No. NAI-54-771; December, 1954.

ntributors

Ackerlind

Ackerlind (A '38—VA '39—SM '52) orn on July 9, 1910, in New York, He received the B.E.E. degree from the Polytechnic Institute of Brooklyn in 1932. In 1934, he received the M.S. in E.E. degree from Columbia University. From 1934 to 1937, he was a research fellow at the Polytechnic Institute of Brooklyn. After obtaining the doctorate in electrical engineering Polytechnic Institute of Brooklyn, oined Hazeltine Electronics, Little N.Y., as an engineer. In 1941, Dr. lind became subsection head at the Research Laboratory and was ed in analytical and experimental igations of direction finders.

joined Northrop Aircraft in 1946 and section head responsible for digital ıter development. In 1949, he became supervisor at the Jet Propulsion atory, California Institute of Techv, and worked on analog computer opment. Since 1953, Dr. Ackerlind een with the Radio Corporation of ica, Los Angeles, Calif., and is now ger of Systems Engineering.

is a member of Eta Kappa Nu and Xi.

❖

S. Bendat

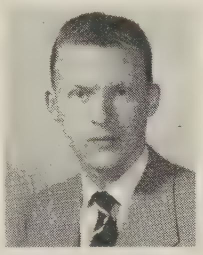
ius S. Bendat was born October 26, in Chicago, Ill. He received the A.B. e in mathematics and physics from the University of California in 1944, and the M.S. degree physics from the California Institute of Technology in 1948. In 1953, he obtained the Ph.D. degree in mathematics from the University of Southern California.

In October, 1955, Dr. Bendat joined aff of the Control Systems Division of Ramo-Wooldridge Corporation, Los es, Calif., where he is now engaged on ıced studies of optimum filter theory ındom phenomena. He is also a lecturer thematics at the University of Southalifornia, specializing in applied mathcs courses.

st experience includes work as an imental physicist on the Manhattan ct at the Radiation Laboratory, ersity of California, from 1942 to 1945; officer in the U. S. Navy during 1946; assistant professor of aerocal engineering at the U.S.C. College eronautics in 1948–1949; research engineer with the Guided Missile Division of Northrop Aircraft Inc., from 1953 to 1955.

Dr. Bendat is a member of Sigma Xi, Phi Beta Kappa, Pi Mu Epsilon, the American Mathematical Society, the Mathematical Association of America, and the Society of Industrial and Applied Mathematics.

❖



J. L. Brown, Jr.

John L. Brown, Jr., was born in Ellenville, N.Y., on March 6, 1925. He received the B.S. degree in mathematics from Ohio University in 1948 after serving in the U. S. Army for three years during World War II. From 1948 to 1951, he held a fellowship in applied mathematics a t Brown University and received the Ph.D. in applied mathematics from that institution in 1953.

In 1951, Dr. Brown joined the staff of the Ordnance Research Laboratory, Pennsylvania State University, as a member of the Theoretical Studies Section. At present, he is an associate professor of engineering research, engaged in applied mathematical research related to the field of underwater acoustics.

Dr. Brown is a member of the American Mathematical Society, Acoustical Society of America, Society for Industrial and Applied Mathematics, Phi Beta Kappa, and Sigma Xi.

❖

For a photograph and biography of L. Lorne Campbell, see page 155 of the December, 1956 issue of IRE Transactions on Information Theory.

❖

D. A. Darling

Donald A. Darling was born in Los Angeles, Calif., on May 4, 1915. He received the B.A. degree from the University of California, L o s Angeles, in 1939 and the Ph.D. degree from the California Institute of Technology in 1947.

Dr. Darling was a research director at the California Institute of Technology from 1942-1945, a teaching fellow from 1944–1947, and a member of the research staff of the Naval Ordnance Testing Station, Inyokern, Calif., during 1945–1946.

He was a research associate, Cornell University, Ithaca, N.Y., 1947–1948; assistant professor, Rutgers University, New Brunswick, N.J., 1948–1949; instructor, University of Michigan, Ann Arbor, Mich., 1949–1950, assistant professor, 1950–1955, and he is now associate professor of mathematics.

He has served as visiting professor at Columbia University, 1952–1953, and the University of Chicago, 1955–1956. He has been a consultant for The RAND Corporation, Santa Monica, Calif., since 1949; for the Operations Research Office, Washington, D.C., since 1952, and for the Engineering Research Institute of the University of Michigan since 1950.

He is a member of the American Mathematical Society, Mathematical Association of America, and Sigma Xi, and a fellow of the Institute of Mathematical Statistics.

❖

J. Dugundji

J. Dugundji was born in New York, N.Y. on August 30, 1919. He received the B.A. degree from New York University in 1940, and in 1942 began four years of service with the Army Air Force. In 1948, he received the Ph.D. degree in mathematics from Massachusetts Institute of Technology. He is presently an associate professor of mathematics at the University of Southern California, and since 1953, has been a mathematical consultant to the Radio Corporation of America in Los Angeles, Calif.

Dr. Dugundji is a member of Phi Beta Kappa, Sigma Xi, and the American Mathematical Society.

❖

For a photograph and biography of Paul E. Green, Jr., see page 97 of the June, 1956 issue of IRE Transactions on Information Theory.

❖

A. Hauptschein

A. Hauptschein (S '47—A '50—M '55) was born on October 31, 1925, in New York, N.Y. He received the B.S. degree in electrical engineering from Pennsylvania State University in June, 1947. In June, 1948 he received the M.S. degree in electrical engineering anp in February, 1957, the professional (E. E.) degree, both from Columbia University.

From 1948 to 1952, he was employed by Airborne Instruments Laboratory as a project engineer in the antenna and special devices section, and worked on the design of communication, navigation, and homing antennas for high speed aircraft and helicopters.

Since 1952, Mr. Hauptschein has been associated with the research division of New York University, department of electrical engineering, in the capacity of engineering scientist and instructor. At New York University he has been concerned with the design of a microwave impedance measuring bridge and is presently engaged in an evaluation study for communication systems.

Mr. Hauptschein is a member of Tau Beta Pi, Eta Kappa Nu, Pi Mu Epsilon, and Sigma Xi.

❖

S. Muroga

Saburo Muroga was born in Numazu, Japan, on March 15, 1925. He graduated from the electrical engineering department of the University of Tokyo in 1947. He was engaged in theoretical research on pulse modulation and narrow-band voice transmission system in the Railway Technical Laboratories from 1947 to 1950 and in the Radio Regulatory Commission from 1950 to 1951.

In 1951, he joined the staff of the Electrical Communication Laboratories of the Nippon Telegraph and Telephone Public Corporation and has been engaged in research of the communication theory and also in construction of a parametronic digital computer. He studied at the research laboratory of electronics of the Massachusetts Institute of Technology in 1953 and at the digital computer laboratory of University of Illinois in 1954.

He is a member of the Institute of Electrical Communication Engineers of Japan and the Physical Society of Japan.

❖

L. S. Schwartz

Leonard S. Schwartz (S '42—A '45—SM '47) was born on May 28, 1914, in Pittsburgh, Pa. He received the B.S. and M.S. degrees in physics from the University of Pittsburgh. While in the military service during World War II, Mr. Schwartz was engaged in radar research and development at the Radiation Laboratory of M.I.T. and later at the Naval Research Laboratory. After the war he remained as a civilian at Naval Research Laboratory until 1947, when he joined the Hazeltine Electronics Corporation to work on radar developments. In 1952, he joined the research division of the college of engineering of New York University where he has been directing projects concerned with applications of communication theory.

Mr. Schwartz is a member of the American Physical Society, the IRE, the AIEE, Sigma Pi Sigma, and Sigma Xi.

❖

A. J. F. Siegert

Arnold J. F. Siegert was born in Dresden, Germany, in 1911. He received the Ph.D. degree at the University of Leipzig, Germany, in 1934. He worked as a Lorentz Foundation fellow in Leiden, Holland, from 1934–1936; as a teaching assistant in the Physics Department, Stanford University, Stanford, Calif., during 1936—1939; as a physicist for the Texas Company from 1939–1942, and the Stanolind Oil and Gas Company, Tulsa, Okla., during 1942.

From 1942 to 1946, he was engaged in war research at the Radiation Laboratory, M.I.T., Cambridge, Mass., and from 1946 to 1947 he was an associate professor in the Physics Department, Syracuse University, Syracuse, N.Y. Since 1947, he has been professor of physics at Northwestern University, Evanston, Ill.

During 1953–1954, Dr. Siegert held a Guggenheim Fellowship and worked at the Institute for Advanced Study, Princeton, N.J. He is a consultant for the Stanolind Oil and Gas Company and The RAND Corporation, Santa Monica, Calif.

He is a fellow of the American Physical Society and a member of the Society of Exploration Geophysicists and the Institute for Mathematical Statistics.

❖

G. C. Sponsler

George C. Sponsler was born December 2, 1927, in Collingswood, N.J. He attended Princeton University and as an undergraduate held an RCA scholarship. He was graduated in 1949 with highest honors, receiving the B.S. degree in engineering. In 1951, he received the A.M. degree and in 1952, the Ph.D. degree, also from Princeton University. He was elected to Phi Beta Kappa and as a graduate student was awarded the Sayr Fellowship and a G. E. Coffin National Fellowship in electronics.

Between terms, he was employed by the Carnegie Institution Department of Terrestial Magnetism, the Johns Hopkins University Applied Physics Laboratory, the RCA Laboratories, and the Brookhaven National Laboratory. During this period he devised and tested equipment for a cosmic ray telescope, mass spectrograph, and an
ionospheric recorder, and performed e
periments with secondary electrons. Aft
receiving his doctor's degree, he w
employed by the Massachusetts Institu
of Technology, Lincoln Laboratory, f
four years, where he worked on the statisti
of radar detection, electron optics, and t
mathematics of systems analysis. I
addition to various mathematical analyse
he devised an automatic electron trajecto
tracer.

At present, Dr. Sponsler is fulfilling
two-year contract as liaison officer f
electronics with the London branch of t
Office of Naval Research.

❖

G. L. Turin

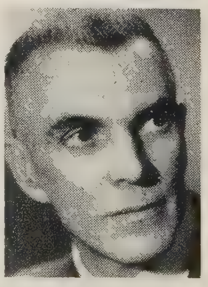
George L. Turin (M '56) was born
New York, N.Y., on January 27, 1930. H
received the B.S. and M.S. degrees from t
Massachusetts Inst
tute of Technolog
in 1952 after co
pleting the cooper
tive course
electrical enginee
ing, with Phil
Corp. as the c
operating compan
In the summer
1952, he was a
M.I.T. Overse
Fellow at Marconi
Wireless Telegra
Co. in England.

From 1952 to 1956, he worked at M.I.T.
Lincoln Laboratory in the field of statistic
communication theory, first as a st
member, and later as a research assista
while completing his doctoral studi
During this latter period, he was also
consultant to the firm of Edgerton, Ge
meshausen and Grier. He received t
Sc.D. degree in electrical engineering fro
M.I.T. in 1956.

Since July, 1956, Dr. Turin has be
engaged in radar research studies at Hugh
Aircraft Co. He also currently teach
part-time at the University of Southe
California.

Dr. Turin is a member of Eta Kap
Nu, Tau Beta Pi, and Sigma Xi.

❖



J. A. Ville

Jean A. Ville was born in 1910 in Ma
seilles, France. He holds the Ph.D. degr
in mathematics and is also a graduate
law (1939). He w
professor of m
chanics at the U
versities of Poiti
and Lyon.

In 1943, he enter
the technical staff
the Société Al
cienne de Constr
tions Mécaniques
Paris for studies
electronics and pr
ability.

Dr. Ville is now
member of the Board of Directors
S.A.C.M., and teaches automatic comp
tation at the Faculty of Sciences of Pa

INFORMATION FOR AUTHORS

Authors are requested to submit editorial correspondence or technical manuscripts to the Publications Chairman for possible publication in the PGIT Transactions. Papers submitted should include a statement as to whether the material has been copyrighted, previously published, or accepted for publication elsewhere.

Papers should be written concisely, keeping to a minimum all introductory and historical material. It is seldom necessary to reproduce in their entirety previously published derivations, where a statement of results, with adequate references, will suffice.

To expedite reviewing procedures, it is requested that authors submit the original and two legible copies of all written and illustrative material. The manuscript should be double-spaced, and the illustrations drawn in India ink on drawing paper or drafting cloth. Each paper should include a carefully written abstract of not more than 200 words. Upon acceptance, papers should be prepared for publication in a manner similar to those intended for the Proceedings of the IRE. Further instructions may be obtained from the Publications Chairman. Material not accepted for publication will be returned.

IRE Transactions on Information Theory is published four times a year, in March, June, September, and December. A minimum of one month must be allowed for review and correction of all accepted manuscripts. In addition, a period of approximately two months is required for the mechanical phases of publication and printing. Therefore, all manuscripts must be submitted three months prior to the respective publication dates. In addition, the IRE Convention Record is published in July, and a bound collection of Information Theory papers delivered at the annual IRE National Convention is mailed gratis to all PGIT members.

All technical manuscripts and editorial correspondence should be addressed to Laurin G. Fischer, Federal Telecommunication Labs., 492 River Road, Nutley, N. J. Local Chapter activities and announcements, as well as other nontechnical news items, should be addressed to Nathan Marchand, Marchand Electronic Labs., Riversville Road, Greenwich, Conn.